Productivity, Wages, and Marriage: The Case of Major League Baseball*

Francesca Cornaglia
f.cornaglia@qmul.ac.uk
Department of Economics
Queen Mary University of London
London, United Kingdom

Naomi E. Feldman
nfeldman@bgu.ac.il
Department of Economics
Ben-Gurion University
Be'er Sheva, Israel

November 9, 2010

Abstract

Using a sample of professional baseball players from 1871 - 2007, this paper aims at analyzing a longstanding empirical observation that married men earn significantly more than their single counterparts holding all else equal. There are numerous conflicting explanations, some of which reflect subtle sample selection problems (that is, men who tend to be successful in the workplace or have high potential wage growth also tend to be successful in attracting a spouse) and some of which are causal (that is, marriage does indeed increase productivity for men). Baseball is a unique case study because it has a long history of statistics collection and numerous direct measurements of productivity. Our results show that the marriage premium also holds for baseball players, where married players earn up to 20% more than those who are not married, even after controlling for selection. The results are generally robust only for players in the top third of the ability distribution and post 1975 when changes in the rules that govern wage contracts allowed for players to be valued closer to their true market price. Nonetheless, there do not appear to be clear differences in productivity between married and nonmarried players. We discuss possible reasons why employers may discriminate in favor of married men.
(*JEL* J31, J44, J70)
Key Words: Marriage Premium, Wage Gap, Productivity, and Baseball

*We thank Kermit Daniel and Sandy Korenman for providing their dissertations, David Yermack for providing contract data, Jerome Adda, Josh Angrist, Oscar Volij, Todd Kaplan, Chris Tyson and seminar particpants at Hebrew University of Jerusalem, LSE, MIT and Tel-Aviv University for comments. We are also indebted to Jim Gates, Library Director of the Hall of Fame for providing unfettered access to the library's archives and resources, Bill Deane and his excellent team of freelance baseball researchers and David Katz, avid White Sox fan. We are grateful to the ESRC for funding [RES-000-22-3267]. PRELIMINARY, PLEASE DO NOT CITE WITHOUT PERMISSION.

1 Introduction

The effect of marriage on wages has been long debated in the economic literature. The main conclusion in standard cross-sectional log wage regressions is that married men are estimated to earn a "marriage premium"–roughly 10 - 40 percent higher wages than their single counterparts. There are a number of proposed explanations for this finding that can be broadly grouped into issues related to endogenous selection and issues related to causal impacts. In particular, selection may be based on unobserved characteristics that are correlated with both marital status and productivity. Additionally, the positive correlation between marriage and wages may be due to reverse causality where men with high wages or high wage growth tend to be more successful in the marriage market. In contrast to this, another line of explanations take the view that marriage has a causal effect on wages. This may be due to employer discrimination (married men are seen as more "stable") or, as many have concluded, productivity differences due to specialization between household and non-household work afforded by marriage. Because men are more free to concentrate on non-household work, they therefore become more productive workers. The marriage premium is of particular interest for analyzing gender-based discrimination in labor markets, as the male marital pay premium accounts for about one-third of estimated gender-based wage discrimination in the United States (Neumark 1988). We investigate the relationship between wages, marriage and productivity using data on professional baseball players. Our analysis will provide evidence as to whether there exists some basis to this observed discrimination once productivity is taken into account.

There are a number of aspects to our research that improve upon previous studies. First and foremost, a notable feature of our analysis is that we use direct objective measures of productivity. Specifically, we consider professional baseball players, making use of data we hand collected from the National Baseball Hall of Fame and Museum data depositories and merged with productivity measures from Sean Lahman's Baseball Archive database. This rich dataset allows us to directly assess whether there is a relationship between marriage and productivity as opposed to only an indirect linkage via wages. The fact that we make use of *objective* productivity measures is, to our knowledge, novel in the analysis of the marriage premium.[1] In addition, access to these direct productivity measures helps to mitigate a number of potential problems that

[1] A few papers have used *subjective* productivity measures, such as supervisor ratings. See Section 2.

have plagued the literature. We discuss these problems more in depth in Section 2, however, to provide one example, potential wives in standard data sets may have more information than the econometrian regarding the future earning potenital of the potential husband and those with high earning potential may be more likely to marry. Thus, what would appear to be marriage's effect on wages is, in fact, the reverse relationship. Our productivity measures likely serve as a sufficient statistic for future earning potential and we can therefore control for this issue in any analysis. Other, secondary advantages to our data is that, while most analyses have been limited to marital status during the years of data collection, we often know the year of marriage even if it occurs long before or after the player's career. This allows us to calculate the number of years married and test whether marriage has a cummulative effect on wage and/or productivity. The underlying idea being that marrying at the beginning of the career allows the spouse to share important occupational passages and experience a higher level of involvement in the spouse's work career. This higher commitment in turn means a higher impact of marriage on the spouse's wages and/or productivity (Crute, 1981). In addition, the nature of our data also allow us to take into account that certain individuals eventually marry even if they do not do so during the years in which we observe their productivity. This allows us to compare two otherwise similar married men who differ only by the number of years married.

For motivation, Panel A of Figure 1 presents a +/- 10 year window of median wage (adjusted for inflation) for players who eventually marry during their careers. On the horizontal axis we have "years of marriage" that is zero is the year of marriage, negative before marriage and positive after marriage. The graph shows that prior to marriage, median wage is relatively flat. Upon marriage (or perhaps slightly before) wages begin to steeply slope upwards. Of course, this graph does not imply causality and, in fact, could a priori be at least partially explained by a number of confounding factors such as age or experience. In fact, Panels B and C of Figure 1 present box plots of experience and age for the same window of time, respectively.[2] We see that, at a minimum, there is a general increasing trend in the median (and also for the mean) in the number of years of experience and age as the number of years married increases. For example, looking at Panel B, players who are three years prior to marriage (years of marriage equal to -3) have, on average, 3.3 years of experience. Players who have been married for three years have,

[2]The shaded rectangle of the box plot identifies the range of the middle 50% of the data (between the 25th and 75th percentiles). The line inside the rectangle represents the median value. The lines coming out of the rectangle extend to 1.5 times the inter-quartile range and the dots identify potential outliers.

on average, 6.8 years of experience. Similarly, Panel C shows that players who are three years prior to marriage are, on average, 25 years old while players who have been married for three years are, on average, 29 years old. To the extent that experience and age positively impact wage, Panel A may simply be capturing that men that earn more simply have more experience and are older. That being said, there is still quite a bit a variation in the distributions of experience and age for each year of marriage, both positive and negative. Thus, while is it clear that experience and age are important controls, it would be premature to suggest that they are the entire story.

As previous studies have found, our results also show that marriage and wages are positively correlated. Married men earn roughly between 15 - 20% more than their single counterparts. However, this results holds primarily in two main subsamples of the data: the top one-third of the ability distribution and post-1975 when strict rules governing contracts were overturned and wages became free to respond to market forces. In contrast to the wage results, across the board, marriage appears to have no statistically signficant effect on productivity using a variety of measurements and subsamples of the data. Thus, while marriage appears to impact wage, its primary mechanism does not appear to be through its impact on average productivity. In fact, controlling for productivity in a wage equation shows that the statitically significant effect of marital status remains, though the point estimate is slightly smaller. We hypothesize that an impact on wages may be due to a number of nontangible aspects of marriage that are not necessarily captured by our direct productivity measures such as stability, leadership skills and popularity that possibly lead employers to discriminate in their favor. We provide some evidence in support of this, where the team level fraction of married players is correlated with ballpark attendance and team wins and where, at the individual level, marriage has a negative effect on the variance of our productivity measures. That is, marriage makes for a more *stable* performance but not necessarily a *better* performance.

2 Literature

The observation that married men earn more than their single counterparts has been well documented using many different datasets, across numerous time periods and countries. There have been two main empirical approaches in the marriage premium literature: studies that make use of cross-sectional data and/or those that make use of panel data (see Ribar (2004) for a review of the methodologies). Generally speaking, cross-sectional results (for example, Bellas (1992), Blau and Beller (1988), Blackburn and Korenman (1994), Chun and Lee (2001), Duncan and Holmlund (1983), Hewitt, Western and Baxter (2002), Hill (1979), Kenny (1983), Korenman and Neumark (1991), Krashinsky (2004), Nakosteen and Zimmer (1987), Schoeni (1995)) have found clear evidence of a marriage premium.[3] Attempts are made to control indirectly for cross-sectional variation in ability but cannot dismiss the interpretation that the results are driven by unobserved individual characteristics and the effect is overstated due to selection into marriage.[4] As such, many of the papers also include fixed-effects panel data analyses that attempt to correct for this bias (for example, Bardasi and Taylor (2005), Cornwell and Rupert (1995, 1997), Datta Gupta, Smith, and Stratton (2005), Duncan and Holmlund (1983), Ginther and Zavodny (2001), Hersch and Stratton (2000), Korenman and Neumark (1991), Krashinsky (2004), Richardson (2000), Stratton (2002), Loughran and Zissimopoulos (2009), Neumark (1988), Rogers and Stratton (2005)). Panel data results have been mixed, some studies find no statistically significant effect of marriage on wages while others find a residual positive effect. These studies generally conclude that there is some causal effect of marriage on wage, whether it is on productivity or merely discrimination is often unresolved.

All of these studies, both cross sectional and panel data, typically include a measure of wage (or log wage) for the dependent variable and a binary indicator for marital status, or some variation of marital status (never married, cohabitation, divorced) or length of marriage along with other demographic controls such as age, education, experience, race, as indirect indicators for time-constant ability where appropriate.[5] Cross-sectional studies have typically estimated a marriage premium ranging between 6% and 35%. While most panel data estimates have confirmed this postive correlation between wage and marital status, some have found the effect to be indistinguishable from zero (e.g. Gray 1997).

We characterize broadly the main explanations in the literature of the marriage premium into issues of selection and issues of causality. Under selection, we have the following explanations: (1) men with high unobserved ability exhibit characteristics that are more likely to be

[3]A number of papers [see, for example, Cohen (2002), Loh (1996), Richardson (2000)] have also considered cohabitation status as separate from never-married and typically find a cohabitation premium that is less than the marriage premium but nontheless positive and signficant. Stratton (2002) also considered cohabitors but found that once taking into unobservable individual effects, the premium disappears.

[4]Krashinsky (2004) and Antonovics and Town (2004) have used first differenced data on twins to account for unobserserved ability. The first study finds that the mariage premium is statistically indistinguishable from zero among twin pairs while the latter finds that the marriage premium remains positive and signficiant.

[5]Loh (1996) found that the association between marriage and wage for men was highly sensitive to observed controls.

found attractive by both employers and potential spouses (for example, stability, industriousness, physical appearance, etc.); (2) married men (or men who are likely to marry) may tend to sort into professions that have higher wages and less non-pecuniary benefits; and (3) reverse causality–men with high wages or wage growth may find themselves facing an improved pool of potential spouses and therefore more likely to marry. Under causality, we have the following explanations: (1) specialization between household and nonhousehold work between the spouses and, relatedly, spousal investment in augmentation skills. In other words, the wife invests in activities that cause the husband to be more productive in the workplace (Becker, 1981); and (2) employer discrimination for a given level of productivity. Employers that exhibit a preference for married workers are not necessarily discriminating against single workers if married workers are more productive. Yet, even when controlling for current productivity, employers may still prefer married workers because they may be more stable, less mobile, exhibit leadership skills, among other reasons.

The first two explanations under selection have been well addressed in the literature. As we mentioned, in order to account for unobserved ability, researchers have used panel data with fixed effects models, under the assumption of time-constant ability. An example is the analysis done by Korenman and Neumark (1991) using data from the National Longitudinal Survey of Young Men (NLSY-M). They find that selection on the basis of fixed unobservable characteristics accounts for less than 20% of the observed wage premium. In a later contribution, Bardasi and Taylor (2005), using data from the British Household Panel (BHPS), show that when moving form OLS to FE the marriage premium falls from 0.09 to 0.02. A zero marriage premium when taking into acount of individual effects is also found in earlier studies such as Cornwell and Rupert (1995 and 1997) and Gray (1997), and in more recent contributions like that by Krashinsky (2004). A number of papers have found evidence of sorting (among these Petersen, Penner and Hogsnes (2006) and Korenman and Neumark (1991)). They have found that the marriage premium disappears once controlling for profession. Consistent with past literature, we are able to estimate our model using standard fixed effects estimation due to the panel nature of our data. Moreover, given that our entire sample is in the same profession, the sorting issue is of significantly less concern in this particular case.[6] The third explanation under selection issues

[6]It is true that the type of man that selects into professional baseball is not necessarily representative of men in general. We discuss generalizations of our results in our conclusion.

has received considerably less attention in the literature. To our knowledge, the sole paper to consider the problem of reverse causality is Korenman (1988). Korenman provides evidence that wages are not positively correlated with future changes in marital status, a fact that makes the reverse causality argument less of a concern. Because we are able to control for productivity, it is unlikely that the potential spouse has more information regarding the future earning potential of the player than the econometrician in this case. Moreover, to the extent that productivity (or changes in producitivity) does not fully explain wage (or changes in wage), conditional on a number of other controls, we are able to make use of institutional details in the setting of contracts that impact wages and wage growth but are arguably uncorrelated with marital status (see Section 7.3).[7]

The causality explanations have, for the most part, received indirect support in the literature. The aformentioned papers that found a residual effect of marriage on wage, after controlling for individual fixed effects and other controls, generally interpret the finding of a statistically significant coefficient on marital status as the causal effect of marriage on wage that arises from specialization.. Attempts have been made to test this causal explanation by controlling for things like hours worked by the wife. The idea is that marriage allows a man to focus on non-household labor while the wife engages in traditional household labor. Evidence is mixed. Many of the papers that have contributed to this literature, for example, Daniel (1995), Gray (1997), Chun and Lee (2001), and Bardasi and Taylor (2005) find a wage penalty associated with wife's labor hours. Hersh and Stratton (2000), on the other hand, using data from the National Survey of Families and Households, find that household specialization does not seem to be responsible for the marriage premium. Along the same line are the results by Loh (1996). He finds no evidence that wives' labor force participation underlies the return to marriage for men. Similar findings are obtained by Jacobsen and Rayack (1996) using the Panel Study of Income Dynamics (1984-1989), and by Hotchkiss and Moore (1999) using the Current Population Survey. Daniel (1993) using the NLSY highlights some racial differences in the marriage premium, in particular he finds that it is inversely related to the wife's hours of work only for white men.

[7] Another take on this issue is that it is less of problem in our particular setting than it would be with more standard panel data sets. For close to the past 40 or so years, baseball players have been extremely high earning relative to the population. Median wages as well as the MLB minimum wage have increased exponentially in the modern period. Thus, the question is whether players see marginal improvements in their spousal applicant pool and probability of marrying as their careers progress and wages increase from already high levels to even higher levels. Or, does the biggest improvement in the applicant pool and the probability of marrying come when expectations of entering MLB pass a certain threshold. We tend to believe the latter but do not entirely dismiss the former argument and therefore address the concerns raised.

There are a number of papers that make some use of productivity measures and are therefore particularly relevant for our study. Korenman and Neumark (1991) use data from a personnel file of a large U.S. manufacturing firm from 1976. What is useful from this data is that is contains supervisor performance ratings that provide a measure of worker productivity aside from the worker's wage. In this paper, the authors attempt to measure productivity, albeit somewhat subjectively, and find that nearly all of the marriage premium (from 23% to 2%) disappears once adding pay grade and performance rating dummies. Mehay and Bowman (2005) use administrative data on male U.S. Naval officers in technical and managerial jobs to explore the effect of marriage on several job performance measures (e.g. promotion outcomes and annual performance reviews). They find that married men receive higher performance ratings and are more likely to be promoted than non married men (the result is robust to selection arising from quit decisions). Similarly, Hellerstein et al (1999) use individual and employer data to estimate marginal productivity differences between different types of workers. They then compare these to estimates of workers' relative wages. They find that married men are significantly more productive than unmarried men and that these differences are reflected in relative wages. Despite using direct meeasures of productivity they are, nonetheless, subjective measures and reflect potential biases of those reporting the measures. It is possible that supervisors simply perceive married men to be more productive workers and therefore give them higher performance ratings or more frequent promotions. The productivity measure we use, alternatively, are objective measures based on exogenous, historical measures of productivity.

3 A Model of the Marriage Premium

> *There was one big glitch: these sorts of calculations could value only past performance. No matter how accurately you value past performance, it was still an uncertain guide to future performance. Johnny Damon (or Terrence Long) might lose a step. Johnny Damon (or Terrence Long) might take to drink or get divorced.*
>
> (*Moneyball: The Art of Winning an Unfair Game*, p. 136)

In this section we sketch a model and provide intution for the effect that marriage has on spousal wage.[8] We assume here that marriage impacts salary through two channels. First, it

[8]Our model is inspired by Daniel (1993) and refer the interested reader to his more detailed model. Without loss of

impacts salary indirectly through its positive causal effect on productivity that occurs because the wife engages in particular actions that impact the productivity of the husband. The main purpose of this involvement is to provide her husband with uncluttered time. For example, a wife may engage in home production such as cooking, cleaning and childcare so that her husband can focus on his career with fewer distractions. She may also provide career advice and moral support or simply allow him extra sleep.

Second, we also allow for marriage to impact wages directly as opposed to indirectly via productivity. These direct influences can take on a number of forms which may lead employers to discriminate in favor of married men. For example, a wife may impact her husband's popularity and visability through public image (for example, hosting formal dinners, participating in public events, charity events, etc.) or marriage may increase a man's stability, reliability (among other characteristics) which in turn make him a better teammate. A professional athlete's career is accompained by numerous formal and informal expectations and therefore not only is the management of the athlete's self-image important, but that of their wives is crucial too. The wife represents her husband to the public, providing a visible link between the worlds of work and family (Crute, 1981).[9] In sum, through these two channels, the wife is able to take actions that make each unit of her husband's time in the market more effective and/or more profitable. All of these wage enhancing activities are subsumed under the heading of "augmentation activities."

Thus, a husband's wage is a function of direct augmentation activities and productivity while productivity is, in turn, a function of indirect augmentation activities and innate ability. Both are also functions of other demographic characteristics such as age and race as well as variables such as experience. We assume further that these variables affect men of different ability levels differently. We can therefore model wages and productivity as follows:

$$S = S(P, \tau, X) \quad \text{and}$$

$$P(\rho, t, X), \tag{1}$$

where S is yearly salary, τ is the direct and t the indirect activities that impact spousal

generality, we consider a marriage premium only for the husband though, it is more accurately described as a marriage premium for the higher earning spouse.

[9]"A wife's look and behavior...can even affect her husband's baseball career. You are part of the package, and if you don't look the part, well, some are going to notice." (Gmelch and San Antonio, 2001).

wages, P represents productivity, and X is a vector of other variables that impact wages and productivity, such as age, race, and experience, among others. Ability is captured by ρ, where higher numbers represent higher innate ability.

We focus on the particular case of our model where wives invest solely in augmentation activities and do not work. In addition, leisure is predetermined for both spouses in order to abstract from the labor-leisure decision.[10] Thus, total available time for the wife (T) is divided between the two augmentation activities. The zero labor hours restriction also has anecdotal, as well as more formal support. The demands of a professional baseball career do not facilitate a stable lifestyle where wives could invest in their own careers. The far majority of wives of MLB players do not work outside of the home as they run the households.[11] Moreover, with rare exception (e.g Marilyn Monroe), MLB players earn wages that are much higher that any wage their wives could earn, which discourages wives' participation in the labor market.[12] [13]

There are a number of interesting implications from this simple model. For instance, suppose that two men have different ability but equal productivity, that is, $\rho_1 > \rho_2$ but $P(\rho_1, t_1, X) = P(\rho_2, t_2, X)$. Under the assuption of monotonicity of $P(.)$, $t_1 < t_2$ and therefore $\tau_1 = T - t_1 > T - t_2 = \tau_2$. In words, conditional on equal productivity, the wives of higher ability men spend less time on indirect and more time on direct augmentation than the wives of lower ability men. As a result, $S(P(\rho_1, t_1, X), T - t_1, X) = S(P(\rho_2, t_2, X), T - t_1, X) > S(P(\rho_1, t_1, X), T - t_2, X)$ by monotinicity of s(.). Another way to think about it is as follows: in the case of differing abilities but equal time spent on indirect augmentation, that is $\rho_1 > \rho_2$ and $t_1 = t_2$, we have $P(\rho_1, t_1, X) > P(\rho_2, t_2, X)$. Provided P(.) is quasiconcave , the marginal impact of an increase in t is decreasing in ability.

[10]The assumption of setting labor hours equal to zero for the wife would also arise endogenously from the model given sufficiently large husband wages relative to wives.

[11]Source: email correspondence with Denise Schmidt, attorney for the Baseball Wives Charitable Foundation (BWCF).

[12]Blau and Kahn (2007) and Wolfram and Leber Herr (2008) present interesting evidence that wives are less likely to participate in the labor force the higher is the husband's wage even taking into account that high earning men tend to be married to high earning women. Inspired by the Wolfram and Leber Herr findings an on-line article noted that "...Men who are in the upper ranks of their profession with stay-at-home-wives earn 30% more than men who are married to women who work. Those men who want to reach the highest rungs of their career and earn the most money often need a stay-at-home wife to take care of all other aspects of their life, including raising a family." "MBA Moms Most Likely to Opt Out," *Bloomberg Business Week*, August 21, 2008.

[13]An anonymous CEO "...allegedly stated that his wife should not work but rather should stay home and run the household, host his parties and mother his children since any wage she would make would essentially be insignificant." [Anonymous CEO, in "If Vikram Pandit is ousted from Citi will his wife Swati divorce him?", Divorce Saloon, The Global 24/7 Divorce and Family Law Blog, Friday July 2nd 2010 http://www.divorcesaloon.com/if-vikram-pandit-is-ousted-from-citi-will-his-wife-swati-divorce-him.]

Families maximize utility subject to standard budget constraints

$$\max_{t} \quad u(C) \qquad \text{subject to}$$

$$\begin{aligned} C - S(P(\rho^m, t^f, X^m), T - t^f, X^m) + Y &\leq 0, \\ \tau^f + t^f &\leq T, \\ C, \tau^f, t^f &\geq 0, \end{aligned} \tag{2}$$

where, in addition to the variables described above, C is consumption, and Y is nonwage income. The indexes m and f represent male and female, respectively.

The first order condition with respect to t is as follows:

$$S_1(P, T - t^f, X) \cdot P_2(\rho^m, t^f, X^m) = S_2(P, T - t^f, X^m) \tag{3}$$

The left hand side of equation (3) reflects the return to indirect augmentation while the right hand reflects the implied return on direct augmentation. For a given value of ρ, the wife equates the marginal value of one one more unit invested in τ_f with the marginal value of one more unit invested in t_f. In this model, both spouses are fully invested in one career. Wives form a work pattern which Papanek (1973, p.90) has labelled the "two person career", characterised by '...a combination of formal and informal institutional demands ... (are) placed on both members of a married couple of whom only the man is employed by the institution.

4 A Primer on Baseball

Professional athletes are a subsample of the population where direct measurements of productivity are often observable. In contrast to other team sports, such as basketball and soccer (football), performance in baseball is directly quantifiable and with a number of measures that are relatively independent of the actions of the player's teammates. Moreover, while there have been changes in the rules over time, relatively speaking, baseball is a fairly stable sport with a long history of uniform player statistics collection. The current typical baseball season is 162 games and runs from early April until early October, followed by the post-season tournament in

October that culminates with the World Series. The regular season is typically divided into 81 "home" games, that is, games played in the team's home stadium and 81 "away" games. There are two main types of players in baseball: pitchers and batters, each with their own productivity measurements.[14] The role of pitchers is to prevent the other team from scoring runs, while the role of batters is score runs for the team. The overall goal in the game is to score more runs than the opposing team.

4.1 Productivity Measures

There are a number of productivity measurements for batters, the simplest of which is the "Batting Average" (BA), which is defined as the number of hits divided by the number of opportunities to bat ("at-bats") in a season. Another conventional measure is the "On-Base-Percentage" (OBP), which takes into consideration a number of ways a batter can get on base (hits, walks and hit by pitch).[15] Third, there is "On-Base plus Slugging" (OPS) which combines the OBP statistic with a measurement of the player's ability to hit for power (a weighted average of the number of bases reached per at-bat). A fourth measure, called "Equivalent Average" (EqA), is meant to capture hitter productivity independent of ballpark and league effects.[16] Most modern-day baseball enthusiasts and commentators consider the latter two statistics to be the most accurate measures of a player's productivity. Table 1, in conjunction with Appendix A, provides exact definitions for each of these measures and, for all of them, higher numbers represent higher productivity. All of these productivity measurements are calculable from the Baseball Archive.

4.2 Wage Setting

Wage setting is notoriously complex in baseball with a number of important changes over the past few decades. In 1975, the so called "reserve clause" was struck down by the courts. The reserve clause, which was standard in all player contracts at this time, stated that upon the contract's expiration, the rights to the player were to be retained by the team with which

[14] Pitchers are often batters as well but they are judged by their pitching and not by their batting performance. While there have been players that have excelled in both roles (for example, Babe Ruth), generally speaking, pitchers tend to be weak batters.

[15] OBP only became an official MLB statistic in 1984, however it is possible to calculate it for all years using the information available in The Baseball Archive.

[16] EqA is nearly impossible for the nonprofessional to calcuate from scratch. A simpler version called REqA (raw EqA) is more easily generated from the data. The main difference is that the raw version is not normalized and does not take into account ballpark and league effects.

he had signed. This meant that practically, even though the player's obligations to the team as well as the team's obligations to the player were terminated (at the end of what was generally a six-year contract), the player was not free to enter into another contract with another team. This effectively gave the team market power over the player. Thus, if a player was not happy with his wage or a trade to a particular team the most he could do was refuse to play. Post-1975, players are generally considered to be valued at closer to their true market values at all stages of their careers.

Figure 2 graphically illustrates the effect of the elimination of the reserve clause. Panel A of Figure 2 breaks down the sample into players with less than six years of experience and greater than or equal to six years. While technically the elimination of the reserve clause directly impacted those players with six or more years of experience, the figure shows that the increase in wages was not limited to only those players. Under the expectation that a player would eventually become a free agent, a player is potentially able to extract economic rents earlier in his career. Panel A shows that wages for all players began to more steeply increase post-1975 and Panel B shows in the normalized version of Panel A that the increase in growth for players with less than six years of experience is even slightly higher than for those players with more than six years of experience. Thus, if marriage has an effect on wages, we would expect that its effect would be stronger post-1975 when wages could more freely respond to market factors.[17]

Recall Figure 1 that showed that wages begin to increase around the time of marriage (not controlling for any other factors). In Figure 3 we break down this result by pre- and post-reserve clause. Figure 3 indeed confirms this original result but shows that it is primarily the post-reserve clause years that are driving the increase. Prior to 1975, wages roughly doubled over the 20 year span of the graph. In contrast, post 1975, wages increased eight fold (and peaked even higher at around six years of marriage). Again, we emphasize that these graphs do not control for anything and the regressions show that once including a number of important controls the effects are not nearly as large. Nonetheless, they provide suggestive evidence that marriage is positively correlated with higher player earnings.

[17]During the first six years in the league, players are under contract (with some exceptions) to a particular team. Beginning in 1974, after three years in the league, a player becomes what is called "arbitration eligible" and can renegotiate wage, presumably for better terms. The best players, called "super-twos" may be eligible after two years.

5 Data

The main database we use comes from the Baseball Archive, an extensive database which is copyrighted by Sean Lahman (http://www.baseball1.com). It contains detailed yearly performance information on players and teams from 1871 through the current season (2007, as of the time of this writing). Since the inception of professional baseball, there have been roughly 16,000 players (and just over 83,000 player-years) that have played in at least one Major League Baseball (MLB) game. Our contribution to the data was the addition of number of variables (though not always available for every player in every year): marital status, year of marriage, accurate data on wages, and race. While these variables are generally publicly available, there is no standard electronic source, and were therefore hand-collected on site for each player using the vast archives of the National Baseball Hall of Fame and Museum (HOF) located in Cooperstown, NY, USA. The main data sources were the National Baseball Library and Archive player questionnaire collection and biographical clippings files, Major League team media guides, *The Sporting News Baseball Register*, 1940 - 1968 and Topps Baseball Cards, 1951 - 1990 (for race data). In addition, these main data sources were supplemented by player contracts, newspaper clippings and internet searches when necessary. Interestingly, obtaining data on players from the early part of the 20th century proved to be no more difficult than more contemporary players and often much easier due to the information available in the questionnaires that were stopped in 1985. Wages for players after 1988 were obtained from *USA Today*, which is regarded to be the most accurate source for more recent player wages. Prior to 1988, wages were not generally collected and made public and were therefore collected from various sources housed at the HOF. In addition, wage data is not at all available prior to 1905. Wages do not include deferred payments and incentive clauses, nor do they include any income earned by endorsements, or other activities that are not included in the player's contract with the team. While we would have liked to collect data on the universe of players, we were limited by money resources and the available time of our freelance researchers. We therefore took a simple random sample[18] of 5,000 players (batters and pitchers) that represented 31,000 player-years and ultimately were able to recover data on marital status and/or year of marriage (roughly 27,500 player-years), wages (roughly 18,600 player-years), and

[18]This is generally the case. We provided the freelance researchers will sequential samples of 1000 players. Two of these random subsamples were restricted to more current years (one post-1948 and another post-1988) in order to collect more observations on black players (for a separate project) and increase the probability of finding wage data as it has been publicly available since the late 1980s.

race (roughly 4,800 players).

Table 1 contains summary statistics of the data. Of the 5,000 players for whom we collected data, there are 3402 batters and 1769 pitchers.[19] Because pitchers (a fielding position) are not generally evaluated according to batting productivity measures, we drop them from the analysis and reserve an analysis of pitchers for future research. We additionally lose observations due to a number of reasons. First, for some players we were able to recover marital status but not wages and vice versa. In addition, there are a few missing entries for the covariates from The Baseball Archive, for example, height, weight, right or left handed. These tend to be slightly concentrated in the early years of the data. Finally, we drop observations where the players switches teams mid season, where we could not find race data, and where a player exits for more than one year (see Section 7.2).

The top panel of Table 1 contains rookie year demographic information. While the average values of our demographic characteristics for the hand-collected sample of batters are fairly similar to the full population of batters, we reject equality of means in standard t-tests with the exception of the age variable, where we do not reject the null. Race cannot, obviously, be compared to the full sample as this data is missing in the population. Looking at the race variables, we see that 80% of batters in our full sample are white, 13% are black, 10% are hispanic, and a final category for all other races represents less than 1% of batters. Note that race categories are not necessarily mutually exclusive.[20]

Next, we utilize a number of variables that capture variants of marital status. Our main variable (*married*) is defined as a binary indicator equal to one if the player is married in year t, zero otherwise. Sixty-nine percent of our sample observations are married player-years. This reflects a combination of observations from players that are married during their entire careers and some who marry during their careers (switches from single to married status or vice versa). More precisely, 35% of players marry prior to beginning or in the first year of their MLB careers, while another 50% marry at some point during. Three percent are single during their entire careers but marry at some point after the career ends and the remaining 12% never marry as of 2007.[21]

[19]Some players perform both roles over their careers, hence, the sum of the two numbers is greater than 5,000.

[20]Until 1947, blacks were not allowed in the league until Jackie Robinson famously crossed over the color line. Blacks reached their peak in the early 1980s at around 27% of players. Today they stand at roughly 10% of all players. Race is notoriously difficult to collect because most data on race is collected by simply looking at pictures of players and/or baseball cards. At times, in particuarly with dark-skinned hispanics or ligher-skinned blacks, it is difficult to determine race. Moreover, it is uncertain with which race the players themselves identify.

[21]The last category is problematic because until a player has died, we cannot say for certain he never married. Thus, a

For players who marry, we also attempted to collect the year of marriage. To be clear, depending on when the player was active, we used a number of different data sources to collect the marital status information. For example, for players who had finished their careers by roughly the mid-1980s, our primary source of information was the questionnaires that were typically filled out by the player after the end of the career (or, sometimes by family members of the player was no longer living) and often provided information on month and year of marriage, sometimes the name of the wife and some detail of the relationship (e.g. high-school sweetheart) if the player married. If there was no information that would suggest that the player latter divorced or was widowed, we assumed that the player remained married from that point on and would fill in his marital status accordingly. Alternatively, marital status information for more recent players was typically found in the media guides and would typically report whether the player was married in a particular MLB season. For these cases, we sometimes do not know year of marriage but when looking up the player for *each* year of his career, we can know whether or not he was married or single in that particular season. In some cases this allowed us to back out year of marriage if the player married during the career. That is, if he is reported as single in years $t-1$ and t and married in year $t+1$ then we would record his year of marriage as $t+1$.[22] However, if a player was always married during the career, there would be no way to back out the year in which he married. We also tried to supplement the questionnaires and media guides with other information in the player's file, such as newspaper clippings. The year of marriage variable is useful because it allows us to generate a variable equal to the number of years leading up to the year of marriage as well as the number of years the player has been married for each year he played professional baseball ($yearsmar$). Thus, if marriage has a cummulative effect on wage and/or productivity, we can exploit variation in $yearsmar$ given that two otherwise identical players are married. We have no information on cohabitation, though its certain that some fraction of our single players cohabit without a formal marriage. To the extent that cohabitors experience some of the benefits of marriage only strengthens the finding of any marriage premium. The second panel also reports wage that is adjusted for inflation (1983 base year). The average income across players is quite

player who has been single prior to, throughout, and after his career as of 2007 is classified as never married. Of course, he may marry during a later year of his career or after his career ends.

[22]A note on timing. If a player married in January through March in year t then we recorded his year of marriage as t. If he married April through December then we recorded his year of marriage as $t+1$. This is account for the fact that contracts are generally established for the MLB season by April. Of course, we also consider lags of marital status in our specifications so overall this particular way of defining marital status is robust to variations.

high at over $456,000. This is primarily driven by the fairly steep increase in wage growth that began to occur in the mid-1970s. The standard deviation in wages has also increased over the years, roughly tripling between 1905 and 2007 as baseball began to see its share of "superstar" players.[23]

The third panel contains information on the productivity measures and the fourth panel contains a number of other important variables for the anlaysis. Similar to the demographic characteristics, we reject equality of means between the population and our full hand-collected sample of batters for each of the productivity measures. The productivity measures in the sample are overall slightly higher than the population as a whole. This problem is exacerbated by the fact that we eventually restrict the actual sample used in the estimation to those players for whom we have at least two observations (due to lagged independent variables and fixed effects estimation). Nearly 30% of all players played in only one MLB season. Thus, these players fall from the estimation sample. Recall that we did some stratification on years after 1948. This would affect variables such as wages and career length that have been trending up over time. Thus, because we sampled more heavily from a time period where career lengths are longer, we are more likely to have a higher average value. We are less concerned about differences in our sample that are due to selection on a factor such as time because it is exogenous. We nonetheless note that if our estimation sample is indirectly selected on ability by the mere fact that we require players with at least two years of data, then we may end up with married and single players coming from different ability distributions. Provided that marital status has a positive effect on performance or, similarly, career length, this would mean that single players are predicted to come from a distribution with a higher average ability, holding career length fixed. This only strengthens our OLS findings and reemphasizes the importance of controlling for unobserved time-constant ability in our empirical analysis.

6 Econometric Specification

Due to the extensive data available from The Baseball Archive, we can follow a large sample of players over the span of their careers. Panel data allows us to hold constant individual-specific

[23]There have always been superstar players in terms of ability but the mega wages the current players earn, even when adjusting for inflation, is a relatively modern phenomemum. Babe Ruth earned the top wage in 1927 at $70,000, by 2007 Jason Giambi, the highest paid player, was earning over $23 million. In 2010, Alex Rodriguez took home the top wage at $33 million.

factors, essentially identifying the effect of marriage on productivity from changes in marital status over a player's career and allows differentiating between the self-selection and causality arguments. Identification in the fixed effects specification will be coming off of the 50 percent of players that switch marital status at least once during their careers while an OLS specification uses variation in marital status across players and time. It is obvious that marital status is not the only factor that affects productivity. Many other factors are well known to affect productivity, such as age, experience, the ballpark, movement among teams, the team manager, just to name a few. In addition, because our data spans well over one hundred years, certain historical events such as World War II, the Korean War, and rule changes that influence our productivity measurements over time should be taken into account. In addition to the aformentioned demographic information, we also include team, fielding position, manager, ballpark and year fixed effects as well as indicator variables that capture major rule changes that may impact productivity and/or wages.

Before proceeding with the main empirical analysis, we briefly mention here an important technical issue that can arise in our analysis primarily because career longevity is correlated with variables like productivity and individual characterstics (the better and more reliable the player, the longer generally is the career and the less likely the player is to attrit from the sample). In particular, if marriage has a causal effect on productivity and low productivity may eventually lead to exit from the database, what begins as a random sample of players becomes nonrandom due to this selective attrition. If this selection were simply correlated with unobserved time-constant ability, then standard fixed effects analysis would provide consistent estimated parameters. However, because attrition is potentially affected by marital status, which is a time varying variable, standard fixed effects estimation may no longer be sufficient to deal with the attrition problem due to correlation between the idiosyncratic error term and an explanatory variable. In Section 7.2 we discuss in depth various approaches we take to test for and potentially deal with this issue.

6.1 The Marriage Premium

Before we turn to addressing a main contribution of our approach, that is, the effect of marital status on productivity, we first estimate the effect of marital status on log wages as others have done in the literature. Our baseline specification is

$$\log(wage)_{iy} = \gamma_0 MAR_{i(y-1)} + x'\gamma_2 + \alpha_i + \tau_t + \pi_p + \beta_b + \delta_y + \mu_m + \varepsilon_{iy} \quad (4)$$

where i and y indicate person and year indexes, respectively. Our main coefficient of interest is γ_0 that captures the mean effect of marital status on log yearly wages. Marital status is lagged by one year reflecting the fact that wages are based on prior performance. There are a number of alternative specifications of MAR that we consider that include additional lags of marital status, years of marriage and various interactions with proxies for ability, as will be further explained. The vector x includes a number of individual characteristics as described in Table 1. These include binary indicators for race (not mutually exclusive), height and weight in rookie year, binary indicators of right and left-handedness (not mutually exclusive), age and its square, experience and its square, lagged number of games played in the season (as a proxy for injuries) and binary indicators for American League, three or more years experience in MLB, and six or more years experience in MLB. Finally, α_i, τ_t, π_p, β_b, δ_y, and μ_m represent individual, team, fielding position, ballpark, year, and team manager fixed effects.[24] The idiosyncratic error term is represented by ε_{iy} and is clustered by player.

As previously noted, our preferred estimation is an unobserved effects model that controls for time invariant individual characteristics, particularly ability. Thus, any residual effect of marital status should reflect its causal impact on wage. In this specification, we include only those control variables that vary nonlinearly over time. These include the squared age and experience terms, binary indicators for American League, three or more years experience in MLB, and six or more years experience

Table 2 presents the marriage premium results. We divide the results into pre- and post-1975 (elimination of the reserve clause) as well as OLS and FE specifications. The first four columns are estimated without the demographic and other time varying controls, while the last four columns do control for these factors. All models do control for team, year, fielding position, ballpark and manager fixed effects. Broadly speaking, the left panel of Table 2 shows that married players earn significantly more than their single counterparts (roughly between 15.6

[24] It is possible to also consider team an outcome variable as better players may switch to better teams. The results are not sensitive to the exclusion of team fixed effects.

and 54.8%, depending on the specification).[25] When moving from OLS to FE, the estimated effect falls, consistent with the idea that at least part of the marriage premium can be explained by time constant ability. Moreover, when moving from pre to post 1975, the effect also becomes larger, consistent with the idea that salaries were more flexible and able to respond to market forces post 1975. Recall, however, that we do not yet include potentially important controls like experience and age. Once adding these controls in the right panel, the marriage premium is generally no longer statistically different from zero (with the exception of column 5, where married players are weakly estimated to earn roughly 3.9% more than single players (p-value of .092). It appears, thus, that marital status does not significantly impact wages once taking into account variables such as age, experience, race, etc.[26]

In Table 3, we further break down the results based upon initial expected ability. We replicate Table 2 but now break the sample down into three roughly equal groups based upon the distrubtion of rookie year plate appearances (what we term "low," "medium" and "high" expected ability). We use plate appearances as a proxy for expected ability and skill - we assume that a player that is expected to perform well will be given more play time, all else equal. Granted, the number of plate appearances in the rookie year is not a perfect measurement of expected ability as, for one example, position in the batting lineup also impacts plate appearances.[27] Using alternative proxies such as the number of rookie year at-bats and rookie year batting average to generate the groups provides overall similar results. The left panel again shows that married players earn significantly more than single players and this is true for each level of ability. In contrast, once we introduce the demographic and other time varying controls we find that the marriage premium disappears except for the high ability group, in which case married men are estimated to earn approximately 17.5% more than single men. This result is statistically significant at a five percent level. In addition, the differences in the estimated coefficients between high ability men and low and medium ability men are statistically significant (p-values of .014 and .046, respectively).

Finally, we want to bring attention to the goodness of fit measures. Across both Tables 2 and 3, the R-squared measures are no less than .80 in the OLS models and .89 in the FE models,

[25]These percentages are approximations that are close to the actual change when the estimated coefficient is fairly small. In the case of such a large estimated coefficient like .548, the actual change is far from the approximation at $100 \cdot (\exp(.548) - 1) = 72.98\%$.

[26]Including additional lags of marital status does not significantly change the resultss.

[27]We also adjust the measure to take into account players that begin mid season.

numbers that are fairly high even for panel data. The two most important controls that contribute to such high R-squared measures are year and experience. An OLS regression of log wage on year dummies and experience has an R-squared of .70 reflecting that these two variables explain much of the variation in wages.

In sum, the marriage premium exists for professional baseball players but is most robust for the highest expected ability group of players as well as for post-reserve clause years. Once controlling for unobserved individual effects, the existance of a marriage premium has typically been interpreted as causal – marriage causes men to earn more.[28] The main explanation has been that men earn more because they are now more productive at work.[29] Despite the numerous papers that have argued for this causal effect, none, to the best of our knowledge has been able to directly test the increased productivity hypothesis. In the next subsection, we directly test the effect of marriage on productivity.

6.2 The Effect of Marriage on Productivity

In Tables 4 and 5 we present estimates of the effect of marriage on productivity that are obtained by reestimating Equation (4), replacing the log of wages with the productivity measures. We focus on two main measures–batting average (BA) and on-base plus slugging (OPS). While OPS and EqA are considered to be the more accurate productivity measures, we are only able to generate the raw, non-normalised version of the latter measure due to data availability. All the productivity measures (as well as variations that use the log of the measure) we described in Section 4.1 present a similar story and results not presented here are available upon request.

In order to obtain accurate measures of productivity, we restrict our sample to players that have a minimum number of plate appearances. The reason for this is that our productivity measures are yearly averages and if a player did not have a sufficient number of plate appearances (or at-bats) during the course of a season, we possibly obtain productivity measures that are in the extremes of the distribution. For example, if a player has only a few plate appearances during a season, it is quite feasible for him to have a batting average of zero (0.000) or even a thousand

[28]Baseball players spend a large fraction of the season "on the road," that is, away from their spouses and families. In addition, marital infidelity is rumored to be common. At this point, we do not take a stand on exactly what aspect of marriage leads to higher wages. To the extent that being on the road and marital infidelity reduce the benefits of marriage, this would bias us against finding any effect. Thus, finding an effect would mean that the true effect is even stronger.

[29]A second explanation is that employers discriminate in favor of married men perhaps because they are view as more stable workers. We will return to this later.

(1.000). Suppose his batting average is 1.000, it would be foolish to suggest that this player has extremely high productivity – in fact, the opposite is more than likely. This is a statistical problem in the sense that we do not have enough observation points to accurately calculate a season mean. There is no set rule as to how many observations we need in order to have an accurate measure of productivity. Of course, the more the better but this comes at the cost of losing observations.[30] As such, we chose a number of cutoffs to test the robustness of our ad hoc restrictions. Restrictions of 20, 50 and 100 plate appearances in a season provide roughly similar results. For brevity, we present only the results based upon the 100 plate appearances restriction and other restrictions are available upon request.

Looking across the columns in the first row of Table 4, we see that marital status is very weakly correlated with batting average – the point estimates are practically zero and imprecisely estimated. For example, in our prefered model of column 8 (FE, post-1975 and including controls), the effect of the lagged value of marital status is an increase in batting average of 0.0007 points – a fairly negligible, not to mention statistically insignificant effect, given that the mean of batting averages for the sample is 0.262 (std dev of 0.089). We arrive at a similar conclusion even when OPS is used as the measure of productivity in the bottom panel of the table. Across the board, the results are not significant at any conventional level. This lack of correlation is also generally true even when breaking the sample down by the ability groups as in Table 5. Granted, there are a few estimated coefficients that are statistically significant at a five or ten percent level (see column 5), but these findings are not generally robust to varying the restrictions on plate appearances or other different productivity measures.

It is interesting to point out that sport researchers and commentators maintain that marriage is a hinderance to performance in elite/professional sports. Because the sport necessitates complete dedication in terms of time, energy and focus, marriage, and all that comes with it, has been viewed as disruptive to the demands of the sport. While most evidence is rather informal, Farrelly and Nettle (2007) use a matched sample of married and single tennis players and find that male tennis players perform significantly worse in the year after their marriage compared to the year before, whereas there is no such effect for unmarried players of the same age. Our data, however, do not consistently support the hypothesis that marriage impacts male productivity.

[30]In order to qualify for league awards, a player needs at least 400 at-bats. This restriction is far too high for our purposes as we simply need enough to claim we have an accurate measure of productivity, i.e., the mean.

6.3 The Direct Effect of Marital Status on Wages

In the previous sections, we have provided evidence for a marriage premium for MLB players with no supporting evidence that the primary mechanism is through marriage's impact on productivity. If marriage indeed has a causal effect on wages an open question remains: by what mechanism is marriage impacting wages?

Before continuing, its important to first establish that marriage does have a direct effect on wages, *controlling for productivity*. As such, we estimate the following equation:

$$\log(wage)_{iy} = \gamma_0 MAR_{i(y-1)} + \gamma_1 PROD_{i(y-1)} + x'\gamma_2 + \alpha_i + \tau_t + \pi_p + \beta_b + \delta_y + \tau_t + \mu_{yt} + \varepsilon_{iy}{}^{31} \quad (5)$$

where $PROD$ is productivity (either BA or OPS) and the remaining variables are defined as before. Tables 6 and 7 report the results. The results show that the lagged value of productivity is highly significiant across all columns suggesting that productivity is clearly an important component of wage determination. The estimated coefficients on the marital status variables that were significant in Tables 2 and 3 remain statistically significant but have fallen slightly in absolute value reflecting the fact that while there is no consistent statistically significant relationship between marital status and productivity, there is some correlation and omitting productivity in Tables 2 and 3 resulted in slightly larger marriage premium estimates (this is true even when restricting to precisely the same sample in the regressions with and without productivity). In particular, the estimated coefficient on the lag of marital status for the high ability group retains its significance at 5% with a small decrease in the magnitude (to .167) despite the inclusion of a highly significant variable such as productivity.

Returning briefly to the model from Section 3, we view these latter results that control for productivity as a more direct estimate of the effect of direct augmentation (τ) whereas the marriage premium results that do not control for productivity combine the two types of augmentation. Taking this in conjunction with the results from Section 6.2 where we find that the indirect augmentation activities that arise from marriage do not appear to consistently and robustly benefit players of any ability level, what remains for the effect of marriage on wages are the direct augmentation activities. While we can only use marriage as a rough proxy for investment in aug-

[31] For clarity we use the same Greek symbols to represent the estimated coefficients as in Equation 4 but they are clearly allowed to be different.

mentation activities, the results from this section can loosely be interpreted as supporting the idea that higher ability men are the only group to benefit from marriage, and they do so through the wife's investment in direct augmentation.

7 Threats to Identification

In this section, we address a number of remaining issues that potentially impact our empirical results.

7.1 Contracts

As alluded to in previous sections, contract setting in baseball is fairly complex. Moreover, historical contract data is, to our knowledge, not available in any public forum. We were, nonetheless, able to obtain three years (1994, 1996 and 1997) of "Joint Exhibit 1," an official document produced annually by Major League Baseball (the sport's governing authority) and the Major League Baseball Players Association (the players' union) pursuant to a collective bargaining agreement. The Joint Exhibit 1 contains authoritative, comprehensive descriptions of contract terms for all players active on August 31 of the prior season. These data contain contract information for players with at least three years of experience and cover nearly all players who were under such contracts from the mid-1990s to 2001 (roughly 1470 contracts). There are a number of interesting aspects of this data to note. First, fully 64% of all contracts are for one year and 90% are for three years or less. From our perspective, this is a positive finding. Short term contracts allow for salary to respond more flexibly to changes in marital status, productivity and other factors.

Once merging this contract data to our dataset, we were able to match nearly 800 contract years for 283 players. A second interesting aspect of the data is based upon a standard two-sided t-test of differences in means, we do not reject the null hypothesis that married and single players have different contract lengths (p-value of 0.39). Again, this is a positive finding because it does not support the hypothesis that married (or single) players have preferences for shorter term contracts with higher salaries as opposed to longer term contracts with lower salaries, all else help equal.[32]

[32]It is plausible that to the extent that married men are more risk averse than single men, we may expect that mar-

Finally, we attempted to reestimate our baseline results using this matched data and restricted to observations that were not locked into multiyear contracts, under the hypothesis that salaries would not be flexible after a contract was set. These generally resulted in regressions with approximately 300 observations and were inconclusive. Even so, under the assumption that the 1990s are rather representative of other decades (at a minimum post-1975), we are more confident that contracts are not severely hampering flexibility in salary setting in our main database nor do there appear to be any obvious differences in preferences in contract setting between married and single players.

7.2 Nonrandom Attrition

Parametric and nonparametric hazard models hazard models confirm that married players have, on average, longer careers than single players (unreported). Moreover, taking arbitrary career lengths such as three, four or five years, we found that in a cross-section, when regressing binary indicators for having a career length of at least three, four or five years on marital status, productivity and other demographics, we found that marital status always had a positive and significant effect. Both of these results confirm that marital status is somehow correlated with career longevity, though, a priori, do not eliminate the possibility that it is simply time-constant unobserved ability that explains the correlation.[33] In order to more precisely test whether attrition is correlated with our dependent variables we took a simple approach based upon Nijman and Verbeek (1992), where a lead of the selection variable is included as an explanatory variable in our fixed effects regressions. The selection variable equals one in years in which the player is observed and zero in the year he leaves the sample. This lead of the selection variable was consistently statistically significant, suggesting correlation between the dependent variable and attrition.

We take two approaches to addressing this issue: sample restrictions on experience and inverse probability weighting (IPW) [see Moffitt, Fitzgerald, and Gottschalk (1999) and Wooldridge (2000)].[34] In the first approach, we cut the sample at various years of experience to test the sen-

ried men would, on average, prefer lower salary and longer contract length, which would only strengthen our finding. Nevertheless, there does not appear to be any support to this hypothesis.

[33]In order to eliminate the mechanical relationship between marital status and longer careers (i.e. its precisely because certain players have longer careers that we observe them getting married), we repeated the test where we checked whether marital status in the first three years of the career affects the probability of having a career that lasts six years or more and we again confirmed the positive and significant effect of marriage.

[34]A third approach to dealing with the nonrandom attrition problem could be the use of median regression. The idea

sitivity of our results to the attition problem. We assume that the attition problem is less severe at lower cutoffs. The first five columns of Table 8 replicate column 8 from Table 7 incrementally restricting from four to eight years of experience (covering about 80% of the sample). Aside from column 1, the results are fairly stable and remain statistically significant at a minimum of a 5% level of significance. High ability married players are estimated to earn roughly between 20 - 22% more than single players.[35] The second approach, IPW, involves two steps. First, for $t = 2, ..., T$, we estimate a probit regression of a binary variable equal to one if the player has not left the sample, zero otherwise, on observables *in the first period* when the sample was chosen randomly.[36] We then calculated fitted probabilities, $\widehat{p}_{it}$ and generate weights equal to $1/\widehat{p}_{it}$ (for $t = 1$, $\widehat{p}_{it}$ = 1 for all i).[37] Wooldridge (2000) shows that IPW provides a consistent, asymptotically normal estimator. Generally speaking, player observations later in the career receive larger weights reflecting the lower probability of these later years being observed, conditional on observables. The final two columns of Table 8 report the IPW results. Again, we replicated columns 7 and 8 from Table 7. The results are consistent with the previous findings. Married men are estimated to earn approximately 17.2% more than single men.

7.3 Reverse Causality

An important issue yet to be fully addressed is that of reverse causality. While a number of papers acknowledge that it is potentially a problem, to our knowledge, Korenman (1988) is the only paper that attempts some sort of formal test. He finds no evidence for reverse causality when regressing current wages on future marital status. We are able to undertake a test similar

here is that we are mostly concerned with correlation of time-varying marital status and exit at the lower end of the distribution. Players with sufficiently high ability may be able to experience negative "shocks" to productivity and not be in danger of exit, whereas this same negative shock to a player with low initial ability may be enough to cause his exit from the sample. Median regression is less impacted by the extremes of the sample and intuitively less impacted by the attrition problem. This approach, however, is proving to be extremely computationally intensive and left for future research.

[35]The result from column 1 indirectly provides a additional robustness test. Recall that most players are locked in contracts for at least the first three years of their career. Thus, restricting the sample to players with four of less years of experience simply does not provide enough time for salaries to respond to changes in the covariates. This is further suported by the lack of significance on the lagged productivity measure.

[36]Because there may be some concern based on Table 1 that our estimation sample is not statistically random, we also calculated the weights using the full population of players and without the marital status or race variables. This had qualitatively little effect on the results.

[37]General attrition is quite complicated and, following the assumptions of the literature, we therefore assume that attrition is an absorbing state. This means that we drop all player observations after the first exit. Players temporarily leave MLB for a number of reasons, for instance due to injury or low performance or any other reason that is unobservable to the econometrician. Because partial or one year "breaks" are quite common, we consider a player to have exited only if he is unobserved in the data for more than one year.

to Korenman (1988) where we can regress current wages or wage growth on future marital status but our test is improved for a number of reasons. First, we can make use of the institutional setting unique to baseball that provides exogenous variation in wage but is arguably uncorrelated with marital status. For a sample of players who are not married by their first year in the league, we can test whether they are more likely to get married after their third year post-1973 when the player becomes arbitration eligible or after their sixth year post-1975 when the reserve clause is no longer binding. In both cases, wages tend to sharply jump up. Second, we can check how productivity growth impacts the probability of marriage. If a player performs well, there may be increased expectations that he will eventually be compensated with a higher wage once he can renegotiate. Thus, while his current wage is not at his full earning potential, high levels of productivity or productivity growth may predict an increased future wage and propensity to marry. The fact that we are able to control for current productivity and its growth greatly improves the ability to test for reverse causality. Table 9 presents the results. The dependent variable is a binary indicator of whether or not the player marries in year $y + 1$. As in prior regressions we include all the binary indicator controls (year, team, etc) and the individual controls of age, experience and their squared terms, the race indicator variables, height, weight and right/left handed. We estimate all the models using standard OLS so that we even allow for time constant unobservables to partially explain the probability of getting married. Column 1 is a reduced-form type approach - we are interested in whether players after their third or sixth years are more likely to marry given that wages tend to sharply increase after these milestone years. The results show that players are no more likely to marry. We subsequently include the lagged value of wages (column 2), their growth rate (column 3) and lagged productivity growth rates (column 4). None of the specifications show that lags of wages or productivity are statistically significant predictors of future marital status, suggesting that our main findings are not driven by reverse causality.

8 Final Remarks

The results from Section 6.3 has established that marriage has a direct effect on earnings of the high ability players even when controlling for productivity. This leads to two questions. First, why do employers appear to discriminate in favor of married players and why are the results robust only for the high ability types? These are not simple questions to answer given currently available data. We expect high ability men to benefit more from direct augmentation (activities that directly impact wages) because the benefits of indirect augmentation (activities that indirectly impact wages through productivity) are relatively small by this level of skill. These are men who begin their careers already in the upper third of the ability distribution, marginal gains in productivity are difficult to achieve and are small. Thus, for high ability men, the wife's role has mainly to do with the direct aspects of augmentation activities, such as public image. In contrast, the most likely augmentation channel for the a low ability player has more to do with ways in which the wife can allow him to improve in his job. This can be done by by allowing him to dedicate more scarce resources, like time, to his work. For these men, characteristics that marriage may build such as stability and reliability play a secondary role to their productivity. It is only once productivity is nearing high levels that men find other ways in which to compete and set themselves apart from others in their profession.

Thus, if employers discriminate in favor of married high ability players, we hypothesize that there must be some added benefit to teams to having such players that we have not fully captured by considering only the productivity measures. While married players are not found to be more productive, what is often important from the team's perspective is "the bottom line," the marginal revenue that married players generate may be higher than single players. This may be due to the image and popularity of a player increasing the fan base or perhaps more subtle benefits to the team that are not captured by batting productivity. Marriage may lead to stability, reliability, maturity and leadership skills that single players of the same ability level are less likely to have. This interpretation is in line with the three factor model of interpersonal trustworthiness (ability, integrity, benevolence) established by Mayer et al. (1995). All three factors of ability, benevolence, and integrity can contribute to trust in a group or organization. Ability is only one of these determinats, and married men are more likely than single men to score higher on the other two dimentions (i.e. integrity and benevolence). These later characteristics, in turn, contribute to greater team success. It is difficult, if not impossible, to know how much revenue one particular player generates for the team but there are a number of variables that should be correlated with the positive aspects of image, stability and leadership skills that we can analyze. We take a number of approaches to gain insight into these issues. First, we look at individual level regressions where we estimate whether married players are more likely to become "all-star" players, where "all-star" is a measure of player popularity and skill, among other factors. We also check whether marital status has any impact on Wins Above Replacement (WAR) - a measure that is meant to capture the value of a player (in terms of wins) to the team.[38] Both of these estimations show no statistically significant effect of marriage on the outcome variable (unreported but available upon request). Next, we check whether marital status has any impact on performance stability. While the results in Section 6.2 showed that marital status has no robust statistically significant effect on the mean value of our productivity measures, perhaps marriage has an impact on how consistently a player performs (the variance). Thus, we look at the effect of marital status on the coefficient of variation of OPS in a +/- three year window around changes in marital status.[39] The results of this approach are reported in the first column of Table 10. The results show that the coefficient of variation of OPS is roughly 6.7% less in the three year window post marriage as compared to the three years prior to marriage and is statistically significant at five percent. We allow for different slopes in the pre and post windows (as captured by the running variables $yearsmar$ and $yearsmar \times post$) and we also control for experience, age, race, year and team dummies.[40] This particular result provides support to the hypothesis in the literature that marriage has a positive effect on characteristics such as stability and consistency. In our data, men do not necessarily become better players upon marriage but show evidence of becoming more consistent, less volatile players. It is perhaps on this basis that they are partially better rewarded. Finally, to be consistent with the main empirical approach, we also replicated the marriage premium and BA/OPS in this local linear regression and confirm the main findings where the post period shows statistically higher earnings post-1975 (but not pre-1975) but no effect on productivity. These results are reported in columns 2 - 5.

Panel B of Table 10 presents the results from team level FE regressions. In columns 2 - 3, the log of total yearly ballpark attendance is regressed on the fraction of married team members, the lagged fraction of wins out of total games, the fraction of home games out of total games, and year,

[38] WAR is the the number of wins a player provides the team above what a team would win were it to replace the player with an average minor league player off the bench.

[39] The coefficient of variation, equal to the variance divided by the mean value of OPS for each player, is calculated in rolling windows of three years. For example, in year of marriage equal to -2 (that is, two years before marriage), the coefficient of variation is calcuated using observations from years -4 to -2. Similarly, in year of marriage equal to -1, the coefficient of variation is calcuated using years -3 to -1. While we want to compare a fixed three year window prior to marriage to the same window post marriage, the value for years of marriage equal to 1 and 2 (calcuated using data from -1 to 1 and 0 to 2, respectively) contain some overlap years with the time period before marriage. We therefore drop these two years and compare the pre window of -2 to 0 to the post window of 3 to 5.

[40] We drop the manager, position and ballpark dummies simply because the sample size is now too small to precisely identify all of the estimated parameters.

ballpark and manager indicator variables. and in columns 4 - 5 the fraction of wins out of total games is regressed on batter and pitcher productivity measures (team level batting and earned run averages), the fraction of home games out of total games, and year, ballpark and manager indicator variables.[41] The results show a positive and statistically significant correlation between the average fraction of married players at the team level and park attendance. Increasing the fraction of married players by ten percent (from a mean value of 68 percent) is associated with approximately 2.0 percent higher yearly attendance. This represents roughly 25,000 additional attendees. In addition, there is also a positive and significant correlation with the number of team wins, albeit the correlation is a bit weaker at ten percent. An increase in the average number of married players by 10 percent is associated with an increase of 0.0022 in average team wins. This amounts to just under 1/2 percent increase given the mean value of average team wins of 0.50. These latter result supports the hypothesis that married players may have positive benefits to teams that manifest themselves in greater team popularity and success and lend some explanation as to why team managers and owners may discriminate in favor of such players.

9 Conclusion

Using a large sample of professional baseball players from 1871 - 2007, this paper aims to further investigate a longstanding finding that married men earn more than otherwise comparable single counterparts. The main contribution of our approach is that we also use direct measures of productivity, allowing us to test the hypothesis that marriage has a causal effect on wages through its impact on productivity. We confirm the existance of a marriage premium even after controlling for demographics such as age and experience and unobservable characteristics that are correlated with selection into marriage and higher wages. Our finding, however, is robust only for men from the top third of the ability distribution, where they earn up to 24% more than similar single players. Despite this, we find no robust statistically significant effect on productivity when using a variety of productivity measures. Moreover, we find that even when controlling for productivity directly in a wage equation, the marriage premium remains positive and statis-

[41] We would also like to include total team budgets as an explanatory variable as its clear that teams with larger budgets can afford to attract the most highly skilled players but historical values of total budgets prior to 1988 are extremely difficult to find. Team fixed effects should capture overall average level differences in budgets across teams.

tically significant, though slightly smaller at approximately 20%. This presents an interesting and open question as to how marriage impacts earnings. We explore a number of mechanisms by which marriage directly impacts wages. We find some evidence that team wins are positively correlated with the fraction of married players. Employers may discriminate in favor of married players because they lead to overall greater team success that is not necessarily captured by the productivity measures.

Because few men are professional athletes, it is natural to question whether the results presented in this paper can be generalized beyond the sports industry. Although professional baseball is a unique occupation, it shares certain features in common with other occupations. Playing professional baseball requires long hours of practice, intense competition and significant travel. As such, we view our project as providing insight into other similarly demanding professions such as CEOs, partners at law firms and other high level corporate executives whose measures of productivity are less straightforward. Moreover, the wife is the closest person to the life of a professional athlete. This is again, however, not unique to professional sport. The wife's accessibility to the husband's work world shares similarities to many of these other professions. We also consider our project to be part of a larger group of papers that use very specific data to analyze basic, yet extremely important labor economics questions. Take, for example, labor supply responses to changes in wages. There have been a number of papers analyzing the labor supply of taxi drivers [Camerer, Babcock, Lowenstein, and Thaler (1997), Farber (2005) and Chou (2000)], stadium vendors [Oettinger (1999)] and bicycle messengers [Fehr and Goette (2007)]. These studies produce results that are convincing in their specific setting and may well be general given sufficient replication in alternative settings. Consequently, we view our project as laying the groundwork for further research, perhaps in other individual sports or demanding professions where more direct productivity measurements are able to be collected by the researcher.

10 Bibliography

1. Antonovics, Kate and Robert Town. 2004. “Are All the Good Men Married: Uncovering the Sources of the Marital Wage Premium.” *American Economic Reivew*, 94(2), 317-321.

2. Bardasi, Elena and Mark Taylor. 2005. “Marriage and Wages.” ISER WP, No. 2005-1.

3. Becker, Gary S. 1981. “Treatise on the family.”Cambridge, MA: Harvard University Press.

4. Bellas, Marcia. 1992. “The Effect of Marital Status and Wives’s Employment on the Salaries of Faculty Men: The (House) Wife Bonus.” *Gender and Society* 6 (December), 609-622.

5. Blau, Francine and Andrea Beller. 1988. “Trends in Earnings Differentials by Gender, 1971-1981.” *Industrial and Labor Relations Review* 41 (4), 513-529.

6. Blau, Francine and Lawrence Kahn. 2007. “Changes in the Labor Supply Behavior of Married Women: 1980-2000,” *Journal of Labor Economics*, 25, 393-438.

7. Blackburn, McKinley and Sanders Korenman. 1994. “The Declining Marital-Status Earnings Differential.” *Journal of Population Economics* 7 (July), 247-270.

8. Camerer, Colin, Linda Babcock, George Loewenstein and Richard Thaler. 1997. “Labor Supply of New York City Cabdrivers: One Day at a Time.” 112(2), 407-441.

9. Chou, Yuan K. 2000. “Testing Alternative Models of Labor Supply: Evidence from Cab Drivers in Singapore.” *The Singapore Economic Review* 47(1), 17-47.

10. Chun, Hyunbae and Injae Lee. 2001. “Why Do Married Men Earn More: Productivity or Marriage Selection.” *Economic Inquiry* 39 (April), 307-319.

11. Cohen, Philip N. 2002. “Cohabitation and the Declining Marriage Premium for Men.” *Work and Occupation* 29 (August), 346-363.

12. Cornwell, Christopher and Peter Rupert. 1995. “Marriage and Earnings.” *Economic Review, Federal Reserve Bank of Cleveland*, Q(IV), 10-20.

13. Cornwell, Christopher and Peter Rupert. 1997.“Unobservable Individual Effects, Marriage and the Earnings of Young Men.” *Economic Inquiry* 35 (April), 285-294.

14. Crute, Beverly Jean. 1981. "Wives of Professional Athletes: an Inquiry into the Impact of Professional Sport on the Home and Family." Ph.D. Dissertation, Department of Sociology, Boston College.

15. Daniel, Kermit. 1993. "Does Marriage Make Workers More Productive?" Ph.D. Dissertation, Department of Economics, University of Chicago.

16. Daniel, Kermit. 1995."The Marriage Premium", in The New Economics of Human Behaviour, Ed. Mariano Tommasi and Kathryn Ierulli, Cambridge University Press.

17. Datta Gupta, Nabanita, Nina Smith and Leslie S. Stratton. 2005. "Is Marriage Poisonous? Are Relationships Taxing? An Analysis of the Male Marital Wage Differential in Denmark." IZA DP No. 1591.

18. Duncan, Greg J. and Bertil Holmlund. 1983. "Was Adam Smith Right After All? Another Test of the Theory of Compensating Wage Differentials." *Journal of Labor Economics*, 1(4), October, 366-379.

19. Farber, Henry S. 2005. "Is Tomorrow Another Day? The Labor Supply of New York Cab Drivers." *Journal of Political Economy*, 113(1), 46-82.

20. Farrelly, Daniel and Daniel Nettle. 2007. "Marriage Affects Male Performance in Competitive Tennis Players." *Journal of Evolutionary Psychology*, 5, 141-148

21. Fehr, Ernst and Lorenz Goette. 2007. "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review*, 97(1) 298-317.

22. Ginther, Donna and Madeline Zavodny. 2001. "Is the Marriage Premium Due to Selection? The effect of Shotgun Weddings on the Return to Marriage." *Journal of Population Economics* 14 (May), 313-328.

23. Gmelch, George and Patricia Mary San Antonio. 2001. "Baseball Wives: Gender and the Work of Baseball." *Journal of Contemporary Ethnography* 32, 335-356.

24. Gray, Jeffrey S. 1997. "The Fall in Men's Return to Marriage: Declining Productivity Effects or Changing Selection?" *Journal of Human Resources* 32 (Summer), 481-504.

25. Hellerstein Judith K., David Neumark, Kenneth R. Troske. 1999. "Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations." *Journal of Labor Economics* 3(July), 409-446.

26. Hersch, Joni and Leslie Stratton. 2000. "Household Specialization and the Male Marriage Wage Premium" *Industrial and Labor Relations Review* 54 (October), 78-94.

27. Hewitt, Belinda, Mark Western and Janeen Baxter. 2002. "Marriage and Money: The Impact of Marriage on Men's and Women's Earnings." Discussion Paper DP-007, The University of Queensland, (July).

28. Hill, Martha. 1979. "The Wage Effects of Marital Status and Children." *The Journal of Human Resources* 14(4), 579-594.

29. Hotchkiss, Julie L. and Robert E. Moore. 1999. "On the Evidence of a Working Spouse Penalty in the Managerial Labor Market." *Industrial and Labor Relations Review* 52, 410-425.

30. Jacobsen, Joyce P. and Wendy L. Ryack. 1996. "Do Men Whose Wives Work Really Earn Less?" *American Economic Review* 86(2), 268-273.

31. Kenny, Lawrence G. 1983. "The Accumalation of Human Capital During Marriage by Males." *Economic Inquiry* 21 (April), 223-231.

32. Korenman, Sanders and David Neumark. 1991. "Does Marriage Really Make Men More Prodctive." *Journal of Human Resources* 26 (Spring), 282-307.

33. Korenman, Sanders. 1988. "Empirical Explorations in the Economics of the Family." Ph.D. Dissertation, Department of Economics, Harvard University.

34. Krashinsky, Harry A. 2004. "Do Marital Status and Computer Usage Really Change the Wage Structure?" *Journal of Human Resources* 39 (3), 774-791.

35. Lahman, Sean. 2009. The Baseball Archive Database. www.baseball1.com.

36. Loh, Eng Seng. 1996. "Productivity Differences and the Marriage Wage Premium for White Males." *Journal of Human Resources* 31 (Summer), 566-589.

37. Loughran, David S. and Julie M. Zissimopoulos. 2009. "Why Wait? The Effect of Marriage and Childbearing on the Wages of Men and Women." *Journal of Human Resources* 44 (2), 326-349.

38. Mayer, R.C., Davis, J. H., and Schoorman, F.D. 1995. "An Integration Model of Organizational Trust." *Academy of Management Review* 20, 709-729.

39. Mehay, Stephen L. and William Bowman. 2005. "Marital Status and Productivity: Evidence from Personnel Data." *Southern Economic Journal* 72 (1), 63-77.

40. Moffitt, Robert and John Fitzgerald, and Peter Gottschalk. 1999. "Sample Attrition in Panel Data: The Role of Selection on Observables." *Annales d'Economie et de Statistique* 55-56, 129-152.

41. Nakosteen, Robert and Michael Zimmer. 1987. "Marital Status and Earnings of Young Men: A Model with Endogenous Selection." *Journal of Human Resources* 22 (Spring), 248-268.

42. Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *Journal of Human Resources* 23 (3), 279-295.

43. Nijman, Theo and Marno Verbeek 1992. "Nonresponse in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function," *Journal of Applied Econometrics* (7) 243-257.

44. Oettinger, Gerald S. 1999. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors." *Journal of Political Economy* 107(2), 360-392.

45. Papanek, Hanna. 1973. "Men, Women, and Work: Reflections on the Two-Person Career." *The American Journal of Sociology* 78 (4), 852-872.

46. Petersen, Trond, Andrew Penner, Geir Hogsnes. 2006. "The Male Marital Wage Premium: Sorting Versus Differential Pay." UC Berkeley WP Series, October, Institute for Research on Labor and Employment.

47. Ribar, David C. 2004. "What Do Social Scientists Know about the Benefits of Marriage? A Review of Quantitative Methodologies." IZA DP No. 998.

48. Richardson, Katarina. 2000. “The Evolution of the Marriage Premium in the Swedish Labor Market 1968-1991.” Unpublished Manuscript, Labor Market Policy Evaluation, Uppsala, Sweden, (June).

49. Rogers, William M. III and Leslie S. Stratton. 2005. “The Male Marital Wage Differential: Race, Training, and Fixed Effects.” IZA DP No. 1747.

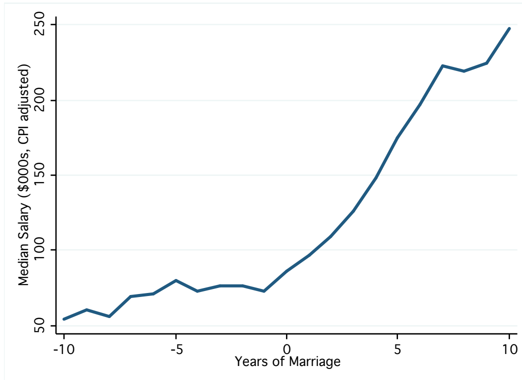
50. Schoeni, Robert F. 1995. “Marital Status and Earnings in Developed Countries.” *Journal of Population Economics* 8 (November), 351-359.

51. Stratton, Leslie S. 2002. “Examining the Wage Differential for Married and Cohabitating Men.” *Economic Inquiry* 40 (April), 199-212.

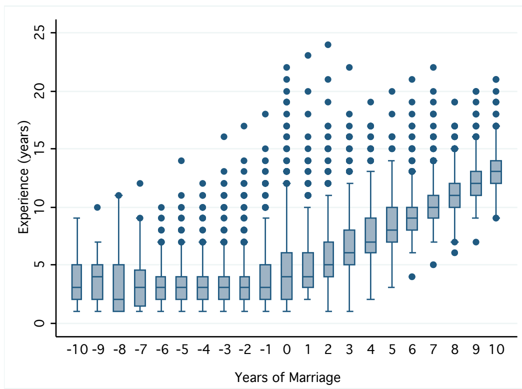
52. Wolfram, Catherine and Jane Leber Herr. 2008. “Opt-Out Patterns Across Careers: Labor Force Participation Rates Among Highly Educated Mothers,” Unpublished Manuscript, Department of Economics, University of California-Berkeley.

53. Wooldridge, Jeffrey. 2000. “Inverse Probability Weighted M-estimators for Sample Selection, Attrition, and Stratification.” *Portuguese Economic Journal* 1(2), 117-139.

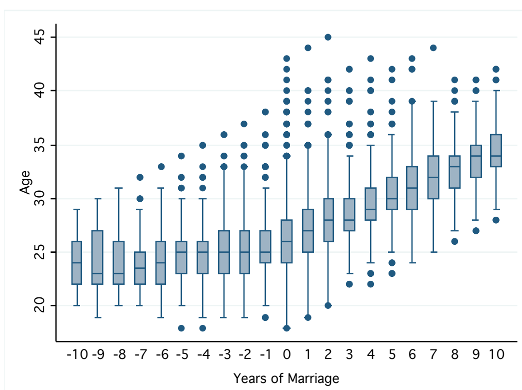
Figure 1: Median Salary, Experience and Age by Years of Marriage



(a) Median Salary



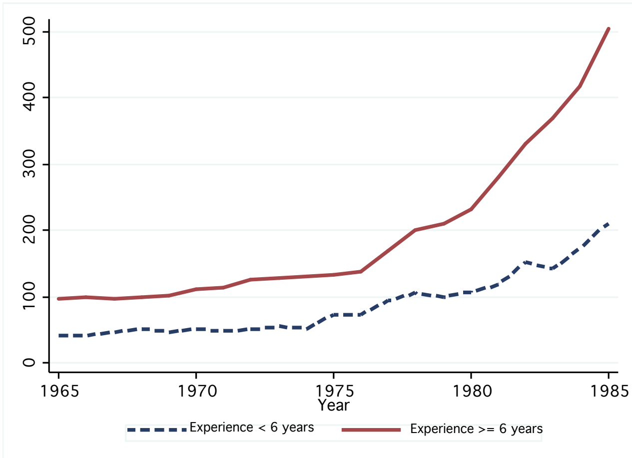
(b) Experience



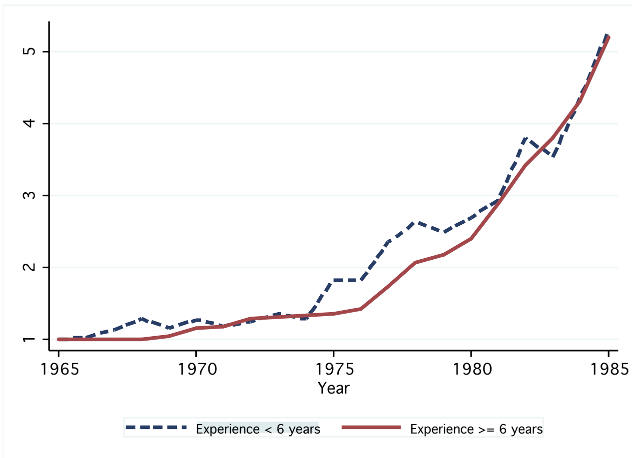
(c) Age

Notes: Author calculations from The Baseball Archive and supplemented with author collected data. Sample restricted to players that marry during the career. Negative numbers on the horizontal axis reflect years leading up to marriage. Box Plots: The shaded rectangle of the box plot identifies the range of the middle 50% of the data (between the 25th and 75th percentiles). The line inside the rectangle represents the median value. The lines coming out of the rectangle extend to 1.5 times the inter-quartile range and the dots identify potential outliers.

Figure 2: Mean Salaries Pre- and Post-Reserve Clause ($000s, CPI adjusted)



(a) Mean Salary



(b) Mean Salary, normalized

Notes: Author calculations from The Baseball Archive and supplemented with author collected data.

Figure 3: Median Salary by Years of Marriage ($000s, CPI adjusted)

Notes: Author calculations from The Baseball Archive and supplemented with author collected data. Sample restricted to players that marry during the career. Negative numbers on the horizontal axis reflect years leading up to marriage.

Table 1: Summary Statistics

Variable	Obs(pop/sample)	Population Mean	Population Std. Dev.	Sample Mean	Sample Std. Dev.	Description
Demographics						
age	9236/3392	24.18	2.84	24.18	2.81	Rookie year age
right	8578/3273	0.70	0.46	0.69	0.46	=1 if the player is right-handed, 0 otherwise
left	8578/3273	0.38	0.49	0.40	0.49	=1 if the player is left-handed, 0 otherwise
white	NA/3386			0.80	0.40	=1 if the player is white, 0 otherwise
black	NA/3386			0.13	0.34	=1 if the player is black, 0 otherwise
hispanic	NA/3386			0.10	0.30	=1 if the player is hispanic, 0 otherwise
otherrace	NA/3386			0.01	0.08	=1 if other race, 0 otherwise
height	7782/2975	71.53	2.37	71.62	2.36	Height in inches
weight	8727/3310	180.17	24.24	181.19	18.63	Weight in pounds (rookie year)
Marital Status and Wages						
married	NA/20204			0.69	0.46	=1 if the player is married, 0 otherwise
yearsmar	NA/15336			3.92	6.07	= Year - Year of Marriage
wages	NA/13872			456.24	1037.46	Yearly Wage ($000s, adjusted for inflation)
Productivity Related Measures						
BA	48960/22902	0.241	0.089	0.247	0.041	Batting Average
OBP	46770/22150	0.307	0.096	0.314	0.048	On-base percentage
OPS	46728/22138	0.651	0.229	0.670	0.124	On-base plus slugging
REqA	33715/12519	0.689	0.220	0.709	0.154	Equivalent Average (raw)
G	49356/23010	72.88	52.26	82.00	50.47	Number of games played
PA	44774/22246	270.53	229.43	306.61	226.08	Plate Appearances
Other Variables						
league	(47960/10430)	0.45	0.50	0.56	0.50	=1 if American League, 0 if National League†
experience	(49357/10430))	5.51	4.40	6.95	4.04	Years in MLB
year	(49357/10430)	1954	38	1978	22	Year (1871 - 2007)‡
position						Fielding position
team						44 unique ids
manager						332 unique ids
ballpark						107 unique ids

1. See Appendix A for further description of the productivity measures
2. Fielding positions include first baseman, second baseman, third baseman, catcher, center field, left field, right field, shortstop. Also includes designated hitter and outfielder.
3. † About 1% of observations are in leagues other than American and National, these are primarily from the early days of baseball and disappeared by the early 1900s. They are included along with the National League in all estimation models.
4. ‡ Wage equations are limited to 1905 - 2007 due to data availability.

Table 2: The Effect of Marital Status on Earnings (The "Marriage Premium")

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$married_{y-1}$	.309	.156	.548	.428	.039	-.007	-.002	.020
	(.031)***	(.030)***	(.056)***	(.064)***	(.022)*	(.024)	(.034)	(.051)
cons.	8.154	6.199	11.552	14.782	4.750	-3.621	8.867	32.260
	(.060)***	(.284)***	(.814)***	(.588)***	(.560)***	(1.197)***	(1.180)***	(2.383)***
Obs	5184	5184	6714	6714	4548	5184	6678	6714
R^2	.638	.912	.554	.826	.867	.947	.801	.888

Dependent variable is equal to $log(salary)_y$. Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience. See Appendix A for the full results.

*10% significance level, **5% significance level, ***1% significance level.

Table 3: The Effect of Marital Status on Earnings by Ability Levels (The "Marriage Premium")

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$married_{y-1} \times ability_1$	.266	.207	.613	.318	.044	-.038	-.052	-.102
	(.066)***	(.066)***	(.089)***	(.104)***	(.032)	(.042)	(.050)	(.085)
$married_{y-1} \times ability_2$	.278	.220	.545	.405	-.003	.031	-.008	-.033
	(.061)***	(.054)***	(.086)***	(.098)***	(.034)	(.039)	(.050)	(.083)
$married_{y-1} \times ability_3$	.334	.094	.489	.545	.069	-.017	.044	.175
	(.046)***	(.045)**	(.097)***	(.107)***	(.035)*	(.038)	(.056)	(.077)**
$ability_1$	7.141		9.348		5.345		8.720	
	(.374)***		(.707)***		(.744)***		(1.237)***	
$ability_2$	7.150		9.566		5.398		8.782	
	(.368)***		(.711)***		(.740)***		(1.244)***	
$ability_3$	7.281		9.978		5.429		8.903	
	(.365)***		(.712)***		(.740)***		(1.233)***	
cons.		6.171		14.738		-3.576		31.568
		(.289)***		(.586)***		(1.173)***		(2.410)***
Obs	5184	5184	6714	6714	4548	5184	6678	6714
R^2	.647	.912	.574	.826	.868	.947	.802	.888

Dependent variable is equal to $log(salary)_y$. Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience..

OLS models are estimated without a constant. R^2 in these models are obtained from the equivalent model estimated with a constant.

*10% significance level, **5% significance level, ***1% significance level.

Table 4: The Effect of Marital Status on Productivity

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BA								
$married_{y-1}$	.003	.003	-.0006	.004	.002	-.002	-.0009	.0007
	(.001)*	(.002)	(.001)	(.002)**	(.002)	(.002)	(.001)	(.002)
$cons$	.451	.223	.270	.261	.103	1.012	.319	.563
	(.055)***	(.056)***	(.021)***	(.019)***	(.050)**	(.125)***	(.052)***	(.087)***
$Obs.$	6954	6954	6091	6091	5995	6953	6058	6091
R^2	.313	.639	.133	.489	.397	.649	.186	.497
OPS								
$married_{y-1}$	.007	.012	-.005	.015	.004	-.004	-.006	.0006
	(.005)	(.006)**	(.005)	(.005)***	(.005)	(.005)	(.004)	(.005)
$cons$	.579	.900	.785	.699	.449	-1.273	.493	1.916
	(.152)***	(.109)***	(.060)***	(.058)***	(.203)**	(.309)***	(.166)***	(.312)***
$Obs.$	6851	6851	6091	6091	5912	6851	6058	6091
R^2	.296	.701	.219	.61	.432	.713	.347	.622

Dependent variable is equal to productivity (BA, OPS). Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience. Sample restricted to observations with at least 100 plate appearances.
*10% significance level, **5% significance level, ***1% significance level.

Table 5: The Effect of Marital Status on Productivity by Ability Levels

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BA								
$married_{y-1} \times ability_1$	.007 (.003)**	-.002 (.005)	.002 (.002)	.003 (.003)	.006 (.003)**	-.008 (.005)*	.0008 (.002)	.0008 (.003)
$married_{y-1} \times ability_2$	.0005 (.003)	-.0004 (.003)	-.0007 (.002)	.003 (.003)	.0002 (.002)	-.006 (.003)*	-.001 (.002)	-.0004 (.003)
$married_{y-1} \times ability_3$	.002 (.002)	.004 (.002)*	-.003 (.002)	.006 (.002)**	.004 (.002)*	.0003 (.002)	-.002 (.002)	.001 (.002)
$ability_1$	.312 (.017)***		.264 (.022)***		.429 (.050)***		.289 (.052)***	
$ability_2$	.318 (.017)***		.267 (.022)***		.434 (.050)***		.290 (.052)***	
$ability_3$	.321 (.017)***		.275 (.022)***		.433 (.050)***		.295 (.052)***	
$cons$		.485 (.022)***		.261 (.019)***		-.110 (.090)		.559 (.087)***
$Obs.$	6412	6412	6078	6078	5530	6412	6045	6078
R^2		.616		.488		.627		.497
OPS								
$married_{y-1} \times ability_1$	.014 (.011)	.009 (.014)	.005 (.009)	.009 (.009)	.013 (.009)	-.012 (.013)	-.005 (.008)	-.0009 (.009)
$married_{y-1} \times ability_2$	-.002 (.009)	.003 (.010)	-.002 (.008)	.009 (.009)	-.005 (.008)	-.015 (.009)	-.006 (.006)	-.005 (.009)
$married_{y-1} \times ability_3$	.011 (.007)	.015 (.007)**	-.014 (.008)*	.023 (.007)***	.015 (.007)**	.002 (.007)	-.008 (.007)	.006 (.007)
$ability_1$	.825 (.057)***		.753 (.061)***		1.150 (.142)***		.402 (.166)**	
$ability_2$	.835 (.056)***		.771 (.061)***		1.160 (.142)***		.410 (.167)**	
$ability_3$	.848 (.056)***		.801 (.061)***		1.162 (.142)***		.422 (.166)**	
$cons$		1.251 (.056)***		.701 (.058)***		-.497 (.255)*		1.890 (.314)***
$Obs.$	6412	6412	6078	6078	5530	6412	6045	6078
R^2	.981	.696	.982	.61	.985	.708	.984	.622

Dependent variable is equal to productivity (BA or OPS). Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience..

OLS models are estimated without a constant. R^2 in these models are obtained from the equivalent model estimated with a constant.

*10% significance level, **5% significance level, ***1% significance level.

Table 6: The Direct Effect of Marriage on Earnings (Controling for Productivity)

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$married_{y-1}$	.254	.125	.524	.412	.036	.001	-.007	.009
	(.028)***	(.029)***	(.051)***	(.068)***	(.023)	(.024)	(.034)	(.052)
OPS_{y-1}	1.641	.816	2.534	1.221	.733	.403	.898	.417
	(.122)***	(.080)***	(.205)***	(.164)***	(.077)***	(.056)***	(.103)***	(.103)***
cons	7.004	12.130	8.377	14.045	4.145	21.346	7.875	31.239
	(.098)***	(.307)***	(.684)***	(.626)***	(.599)***	(1.161)***	(1.354)***	(2.732)***
Obs.	3736	3736	5476	5476	3310	3736	5446	5476
R^2	.733	.935	.617	.834	.886	.958	.811	.895

Dependent variable is equal to $log(salary)_y$. Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience.
*10% significance level, **5% significance level, ***1% significance level.

Table 7: The Direct Effect of Marriage on Earnings by Initial Ability (Controlling for Productivity)

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$married_{y-1} \times ability_1$	.273	.187	.519	.254	.057	.005	-.054	-.101
	(.059)***	(.082)**	(.092)***	(.123)**	(.035)	(.057)	(.058)	(.099)
$married_{y-1} \times ability_2$	.233	.154	.553	.395	-.012	.026	-.018	-.057
	(.060)***	(.060)**	(.080)***	(.101)***	(.036)	(.042)	(.052)	(.090)
$married_{y-1} \times ability_3$	.340	.116	.544	.533	.068	.0006	.048	.167
	(.040)***	(.039)***	(.087)***	(.102)***	(.037)*	(.035)	(.055)	(.073)**
$ability_1$	6.438		9.484				-.100	
	(.286)***		(.670)***				(.063)	
$ability_2$	6.455		9.598		.063			
	(.285)***		(.675)***		(.040)			
$ability_3$	6.492		9.885		.102		.107	
	(.281)***		(.674)***		(.041)**		(.062)*	
OPS_{y-1}	1.621	.781	2.491	1.294	.691	.357	.914	.414
	(.130)***	(.084)***	(.199)***	(.172)***	(.077)***	(.055)***	(.107)***	(.107)***
cons		12.043		14.239		23.048		29.070
		(.144)***		(.656)***		(1.346)***		(3.045)***
Obs.	3736	3736	5476	5476	3310	3736	5446	5476
R^2		.937		.838		.962		.895

Dependent variable is equal to $log(salary)_y$. Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience. OLS models are estimated without a constant. R^2 in these models are obtained from the equivalent model estimated with a constant.
Full results are avilable in Appendix B.
*10% significance level, **5% significance level, ***1% significance level.

Table 8: Endogenous Attrition: Robustness Tests

	Exper. ≤ 4yrs	*Exper.* ≤ 5yrs	*Exper.* ≤ 6yrs	*Exper.* ≤ 7yrs	*Exper.* ≤ 8yrs	OLS-IPW	FE-IPW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$married_{y-1} \times ability_1$	-.181	-.163	-.143	-.105	-.067	-.036	-.005
	(.185)	(.140)	(.112)	(.101)	(.098)	(.058)	(.112)
$married_{y-1} \times ability_2$	.007	-.037	-.050	-.044	-.083	-.044	-.047
	(.117)	(.096)	(.083)	(.081)	(.079)	(.058)	(.099)
$married_{y-1} \times ability_3$	.100	.207	.224	.192	.198	.044	.172
	(.138)	(.098)**	(.084)***	(.075)**	(.072)***	(.058)	(.083)**
$ability_1$						7.988	
						(1.369)***	
$ability_2$						8.098	
						(1.378)***	
$ability_3$						8.191	
						(1.362)***	
OPS_{y-1}	.073	.175	.191	.282	.322	1.056	.479
	(.180)	(.145)	(.123)	(.116)**	(.114)***	(.125)***	(.118)***
$cons$	19.418	36.867	40.113	36.005	41.409		27.232
	(6.003)***	(6.413)***	(5.012)***	(4.185)***	(3.617)***		(3.175)***
$Obs.$	1612	2155	2656	3138	3565	5446	5476
R^2	.954	.939	.933	.929	.924	.998	.891

Dependent variable is equal to $log(salary)_y$. Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience. OLS models are estimated without a constant. R^2 in these models are obtained from the equivalent model estimated with a constant.
*10% significance level, **5% significance level, ***1% significance level.

Table 9: The Effect of Earnings on Future Marital Status (Test of Reverse Causality)

	(1)	(2)	(3)	(4)
$experience_{3yrs+}$	.011			
	(.018)			
$experience_{6yrs+}$	.011			
	(.015)			
$log(salary)_y$		.0003		
		(.005)		
$\Delta log(salary)_y$			.008	
			(.010)	
$\Delta log(OPS)_y$				-.004
				(.019)
$cons$	-.578	-.210	-.196	-.863
	(.240)**	(.349)	(.298)	(.367)**
$Obs.$	4855	4172	3496	4093
R^2	.094	.105	.118	.100

Dependent variable is equal to one if year of marriage occurs in year $y + 1$, zero otherwise. All models are estimated by OLS with robust standard errors, clustered on player id in parentheses. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, and American League dummy.
*10% significance level, **5% significance level, ***1% significance level.

Table 10: The Effect of Marital Status on Performance Stability, Ballpark Attendance and Team Wins

Panel A

	Coef. of Variation	$log(salary)_{1975+}$	$log(salary)_{1975-}$	BA	OPS
	(1)	(2)	(3)	(4)	(5)
$yearsmar$	-.014 (.009)	.148 (.074)**	.072 (.042)*	.002 (.003)	-.001 (.009)
$yearsmar \times post$	.042 (.013)***	-.333 (.091)***	-.058 (.052)	-.005 (.006)	-.014 (.015)
$post$	-.067 (.033)**	.595 (.199)***	.017 (.149)	-.0006 (.015)	-.005 (.038)
$cons$	-.130 (.172)	11.371 (.969)***	7.266 (.646)***	.241 (.051)***	.932 (.152)***
$Obs.$	598	555	405	598	598
R^2	.179	.714	.813	.175	.262

Panel B

	log(attend) OLS	*log(attend)* FE	*wins* OLS	*wins* FE
$married$	.177 (.071)**	.203 (.079)**	.023 (.013)*	.022 (.013)*
$homegames$	4.042 (.904)***	3.325 (.713)***	.276 (.128)**	.266 (.130)**
$wins_{y-1}$	1.597 (.150)***	1.134 (.144)***		
BA			3.631 (.144)***	3.626 (.149)***
ERA			-.107 (.002)***	-.107 (.003)***
$cons$	10.009 (.500)***	11.820 (.397)***	-.253 (.073)***	-.207 (.069)***
$Obs.$	2009	2009	2136	2136
R^2	.897	.936	.853	.854

Panel A regressions are at the player-year level. The variable $yearsmar$ captures the pre-marriage trend and $yearsmar \times post$ captures the post-marriage trend in the dependent variable. The effect of marriage is captured by the $post$ variable. All regressions control for experience, age, race, league and team and year indicator variables. Except for column 3, all regressions restrict to post-1975. Robust tandard errors, in parentheses, are clustered by player.
Panel B regressions are at the team-ballpark level in columns 2 - 3 and at the team level in columns 4 - 5. The dependent variable in columns 2 - 3 is equal to the log of ballpark attendance, in columns 4 - 5 it is the number of wins divided by games played ($wins$). The variable $married$ is the team level fraction of married players. $homegames$ is the fraction of homegames divided by total games. BA is the team level batting average over all batters. ERA is the team level earned runs average over all pitchers. Robust standard errors, in parentheses, are clustered by team-ballpark in columns 2 -3 and team in columns 4 - 5. All regressions control for year effects and team-ballpark (FE only).
*10% significance level, **5% significance level, ***1% significance level.

11 Appendix

A Glossary of Terms

1. PA (plate appearances): At bats (AB) + some of the scenarios excluded from at bats such as base on balls (BB), hit by pitch (HBP), sacrifice (SF) + times reached on defensive interference.

2. AB (at-bats): Batting appearances, not including bases on balls, hit by pitch, sacrifices, interference, or obstruction.

3. BB (base on balls): A base on balls (BB) is credited to a batter and against a pitcher in baseball statistics when a batter receives four pitches that the umpire calls balls.

4. HBP (hit by pitch): Is a batter or his equipment (other than his bat) being hit in some part of his body by a pitch from the pitcher.

5. SF (sacrifice fly): Is a batted ball that satisfies four criteria: 1. There are fewer than two outs when the ball is hit; 2. The ball is hit to the outfield; 3. The batter is out because an outfielder or an infielder running in the outfield catches the ball (or would have been out if not for an error); 4. A runner who is already on base scores on the play.

6. SH (sacrifice hit): Is the act of deliberately bunting the ball in a manner that allows a runner on base to advance to another base.

7. H (hits): A hit, also called a base hit, is credited to a batter when the batter safely reaches first base after hitting the ball into fair territory, without the benefit of an error or a fielder's choice.

8. TB (total bases): The number of bases a player has gained with hits, i.e., the sum of his hits weighted by 1 for a single, 2 for a double, 3 for a triple and 4 for a home run. Only bases attained from hits count toward this total.

9. slugging percentage: Is a popular measure of the power of a hitter. It is calculated as total bases divided by at-bats.

10. BA (batting average): The ratio of hits (H) to at bats (AB).

11. *OBP* (on base percentage): Is a measure of how often a batter reaches base for any reason other than a fielding error, fielder's choice, dropped/uncaught third strike, fielder's obstruction, or catcher's interference. Is determinard by the following formula: (H + BB + HBP)/(AB + BB + HBP + SF).

12. *OPS* (on base plus slugging): OBP plus slugging average.

13. *REqA* (equivalent average): Is a metric intended to express the production of hitters in a context independent of park and league effects. It is give by the following formula: (H + TB + 1.5(BB+HBP) + SB + SH + SF)/(AB + BB +HBP +SH + SF + CS + SB/3).

14. *ERA* (earned run average): Is the mean of earned runs given up by a pitcher per nine innings pitched. It is determined by dividing the number of earned runs allowed by the number of innings pitched and multiplying by nine.

15. All-Star: The Major League Baseball All-Star Game is an annual exhibition baseball game between the best players from the National League and the American League.

B Expanded Results (Table 7)

Table 11: The Direct Effect of Marriage on Earnings by Initial Ability (Controlling for Productivity)

	Without controls				With Controls			
	Pre-1975		Post-1975		Pre-1975		Post-1975	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$married_{y-1} \times ability_1$	.273 (.059)***	.187 (.082)**	.519 (.092)***	.254 (.123)**	.057 (.035)	.005 (.057)	-.054 (.058)	-.101 (.099)
$married_{y-1} \times ability_2$	.233 (.060)***	.154 (.060)**	.553 (.080)***	.395 (.101)***	-.012 (.036)	.026 (.042)	-.018 (.052)	-.057 (.090)
$married_{y-1} \times ability_3$	.340 (.040)***	.116 (.039)***	.544 (.087)***	.533 (.102)***	.068 (.037)*	.0006 (.035)	.048 (.055)	.167 (.073)**
$ability_1$	6.438 (.286)***		9.484 (.670)***				-.100 (.063)	
$ability_2$	6.455 (.285)***		9.598 (.675)***		.063 (.040)			
$ability_3$	6.492 (.281)***		9.885 (.674)***		.102 (.041)**		.107 (.062)*	
age					.170 (.037)***	-.004 (.0007)***	-.009 (.075)	-.007 (.002)***
age^2					-.003 (.0007)***	-.004 (.0007)***	-.0009 (.001)	-.007 (.002)***
$experience$					.140 (.015)***	-.003 (.0007)***	.303 (.031)***	-.007 (.002)***
$experience^2$					-.003 (.0009)***	-.003 (.0007)***	-.009 (.001)***	-.007 (.002)***
$games_{y-1}$					.004 (.0002)***	.002 (.0001)***	.007 (.0003)***	.004 (.0003)***
$Amer.\ League$					4.086 (.783)***	-.545 (.396)	9.183 (1.401)***	-1.059 (.679)
$left$					.060 (.062)		.014 (.048)	
$right$					.101 (.063)		-.044 (.055)	
$black$					.157 (.051)***		.120 (.072)*	
$white$					.111 (.060)*		.019 (.071)	
$hispanic$					.026 (.060)		.041 (.063)	
$other\ race$							.397 (.379)	
$height$					-.009 (.008)		.009 (.011)	
$weight$					.003 (.001)***		.004 (.001)***	
$experience_{3yrs+}$					-.001 (.021)	.073 (.019)***	.284 (.050)***	.401 (.050)***
$experience_{6yrs+}$					.013 (.021)	.017 (.019)	.200 (.041)***	.202 (.041)***
OPS_{y-1}	1.621 (.130)***	.781 (.084)***	2.491 (.199)***	1.294 (.172)***	.691 (.077)***	.357 (.055)***	.914 (.107)***	.414 (.107)***
$cons$		12.043 (.144)***		14.239 (.656)***		23.048 (1.346)***		29.070 (3.045)***
$Obs.$	3736	3736	5476	5476	3310	3736	5446	5476
R^2		.937		.838		.962		.895

Dependent variable is equal to $log(salary)_y$. Robust standard errors, clustered on player id in parentheses. All models control for team, position, manager, ballpark and year effects. OLS controls include: age and its square, experience and its square, race dummies, height, weight, left/right handed, lagged games played, American League dummy and indicators for more than three and more than six years of experience. FE controls include: age and experience squared, lagged games played, American League dummy and indicators for more than three and more than six years of experience. OLS models are estimated without a constant. R^2 in these models are obtained from the equivalent model estimated with a constant.
*10% significance level, **5% significance level, ***1% significance level.