Racial Discrimination or Statistical Discrimination? MLB Rookie Card Values and Performance Uncertainty

Gregory S. Burge* and Arthur Zillante†

Abstract: This paper considers racial discrimination in the market for Major League Baseball rookie cards. Merging data from four sources to form a panel containing 6,026 cards released between 1986 and 1993, we explore the determinants of card values in certain and uncertain environments. Our results indicate black players carry a 15 percent premium at initial release, that race effects are not present after playing careers are finished, and that given comparable career accomplishments, prices for black players decline more over time than white players. Collectively, this suggests statistical discrimination influences investor behavior in the baseball card market, which we note is directly linked with a market (MLB labor contracts) where other research has established race effects are present. None of our estimations reveal significant differences between RC prices for white and Hispanic players.

(*JEL* Classification: D81, J15, J71, L83)

* University of Oklahoma, 308 Cate Center Drive, Norman, OK 73019 (e-mail gburge@ou.edu)
† University of North Carolina at Charlotte, 9201 University Center Blvd., Charlotte, NC 28223 (e-mail azillant@uncc.edu)

Preliminary Draft; All comments are welcome.

I. Introduction

Becker (1971) identified the three possible sources of persistent race based wage differentials as discrimination by employers, discrimination by fellow workers, and discrimination by customers. Because employer discrimination is unlikely to survive in competitive labor markets and household mobility would tend to mitigate worker discrimination over time, he argued customer based discrimination is likely to be the most insidious of the trio. Since then, customer based discrimination has been shown to influence the racial composition of occupations (Bergmann, 1974), wage/salary compensation (Butler, 1982; Neumark, 1999), employment status (Holzer and Ihlanfeldt, 1998; Bertrand and Mullainathan, 2004), and housing opportunities (Yinger, 1986) among other outcomes.

Empirical studies in this literature face two significant obstacles. First, there are frequently concerns that the observable data may not be sufficiently detailed to rule out the possibility that race correlates with unobservable factors that may influence the outcome of interest. In this case, statistical analysis may over or underestimate the role of discrimination. Second, when observing an outcome such as hiring, wages, or promotions, the common claim that employer and worker driven discrimination should be small compared to customer based discrimination is not sufficient to attribute observed race gaps entirely to customer discrimination.

We consider customer based discrimination by examining a market that effectively removes both of these concerns – the market for collectable Major League Baseball (MLB) cards. A unique aspect of the baseball card market is that customers buy cards from local stores, rather than from the players or teams themselves. As such, there is no way for employer or coworker discrimination to *directly* influence card values. Also, player performance measures for MLB players are remarkably detailed, such that concerns related to omitted variable bias are mitigated,

if not removed entirely. Finally, besides overcoming these two common obstacles, we exploit a unique aspect of the baseball card market that generates a natural experiment concerning the presence of information uncertainty over player performance, providing the first study to examine baseball card values when they are initially released.

Our work extends the literature on customer discrimination by examining how race influences card values in two distinct investment settings, where uncertainty over the eventual realization of career achievements plays prominent and inconsequential roles. Specifically, we use a novel hand collected data set that contains each card's first reported price in the industry standard monthly price guide (Beckett Baseball Card Monthly, hereafter Beckett) for all rookie cards (RCs) released by MLB licensed card companies during 1986-1993. We compare the effects of race in this uncertain environment, where we argue the potential for statistical discrimination is strong, to the effect of race on card values 17 years later, when career achievements are largely certain and the potential for statistical discrimination is gone, such that any residual effects of race would have to be attributed to racial discrimination on the part of collectors.

Besides our connection to the customer discrimination literature, we add to the literature considering the role of race in professional sports. For example, research has shown that white players earned higher salaries (Pascal and Rapping, 1972; Gwartney and Haworth, 1974), played higher profile positions (Curtis and Loy, 1978), and were more quickly promoted to the majors (Bellemore, 2001) than similarly qualified black and Hispanic players. Recent evidence even suggests major league baseball umpires are more likely to call strikes for pitchers who share their own race than they are for pitchers who do not (Parsons et al., 2011).

Our empirical results indicate race influences the determination of RC values in the initial release environment, where performance uncertainty is prominent. We find RC values for black

players carry a 15 to 19 percent premium over white players, depending on model specification and the group of players considered. We supplement this finding with a set of stylized facts regarding the statistical distributions of player race, player performance, initial RC values, and final RC values, which illustrates how statistical discrimination could produce this pattern of effects. The ability to empirically disentangle statistical discrimination effects from pure racial discrimination has proven useful in other studies of labor market outcomes (Holzer, Raphael, and Stoll, 2006) as well as in the baseball card marketplace (List, 2004). Also, previous work has shown Hispanic and black players carry a higher likelihood of attaining All-Star designation than white players (Depken and Ford, 2006). However, our study provides the first result of this nature for the baseball card market.

The effect of race on RC values reverses direction in models that consider the change in the value of RCs from the time of initial release to the end of player's careers. We find the change in price over time, controlling for the player's career performance, declines at a greater rate for black players than for white or Hispanic players. Given these countervailing effects, it is not surprising that models explaining final prices do not find significant race effects. The estimated price effects for Hispanic players are always smaller than those estimated for black players, and never achieve statistical significance. Collectively, our findings do not imply collectors of baseball cards display purely discriminatory behavior, at least to a level where it affects the marginal purchase of the cards examined, but instead that their willingness to pay for cards in uncertain environments is likely influenced by statistical discrimination, driven by the correlation between race and the likelihood of attaining the highest levels of career achievements.

Specifically, we show that, while hitters of all classifications show nearly identical levels of productivity when measured on a per-season basis, black and Hispanic hitters have significantly

longer careers and achieve higher career performance totals. Per season and career performances for pitchers does not display any significant correlation with race, although we do find evidence that black pitchers are more likely to fill the role of relief pitchers, whereas white and Hispanic pitchers are more likely to be used as starting pitchers. For this reason, white and Hispanic pitchers accumulate more innings pitched, but pitch in fewer games. Using simple comparisons that do not control for career performance, we still observe meaningful differences in price trends over time are found across the three racial classifications, as the average RC value for white players falls over time, whereas values for Hispanic and black players rise.

In the next section we examine the literature considering discrimination within the market for baseball cards. We then describe our four data sources in Section III. Section IV documents the stylized facts regarding race and MLB performances. Our empirical methodology and results are reported in Sections V and VI. Section VII discusses our work and its implications.

II. Customer Discrimination and Baseball Cards

The literature considering customer discrimination and baseball cards begins with the work of Nardinelli and Simon (1990). They examine the values of 567 cards from 1970 Topps using Beckett's 1989 annual price guide. Their explanation is worth noting because their choice to use values of retired players and to focus on all the cards in a single set rather than accumulating a larger number of cards across multiple sets heavily influenced subsequent studies:

> "We chose the year 1970 to avoid having to account for differential information about young and old players. There is a brisk speculative demand for the cards of young players because dealers and collectors attempt to forecast the future superstars of the game. A young, but promising, player will therefore often sell for more than a better,

> but older, player who has already established a level of performance below superstar level. By choosing 1970 cards, we have no young players as of 1989." (pg. 579)

In estimations that control for career achievements, they find evidence for discrimination against Hispanic hitters and black pitchers. Price reductions for black hitters and Hispanic pitchers are of similar magnitudes, but narrowly fail to achieve statistical significance. On average, they find white hitters sell for 10 percent more than non-white hitters, while white pitchers carry a 13 percent premium.

Subsequent studies have found either similar premium for white players or have produced no evidence of discrimination. Andersen and La Croix (1991), Gabriel, Johnson, and Stanton (1995), Fort and Gill (2000), and Burnett and VanScyoc (2004, 2012) all find evidence of discrimination favoring white players; Regoli (1991), Gabriel, Johnson, and Stanton (1995, 1999), McGarrity, Palmer, and Poitras (1999), and Burnett and VanScyoc (2012) find no evidence of discrimination, although most of these papers note that their results depend on the methods and data used for analysis. While a description of the most relevant results to our study follows, Table 1 first summarizes the findings of the previously published studies of baseball cards values (that we are aware of) by listing the card release years analyzed, the year Beckett price data came from, and the findings of each regarding race variables. Note that some papers have multiple listings, as more than one data set was analyzed.

Andersen and La Croix (1991) and Fort and Gill (2000) include a dummy variable for RCs as a control, making an important step towards making more accurate value comparisons. Both studies find that RC designation positively affect prices, though the positive effect in Andersen and La Croix only registers once a player has attained a threshold level of career

accomplishments. However, as the focus of these papers is not directly on the RC market, the only conclusions made are about RC values relative to other card values in a set.

McGarrity, Palmer, and Poitras (1999) use 1974 Topps cards to argue evidence for racial discrimination is closely tied with the Tobit modeling choice. They find that when other models are used, the premium associated with white players dissipates. Their choice of 1974 Topps interacts uniquely with an important aspect of the market shown by the two previously discussed papers – the premium paid for RCs. A 1974 Topps Dave Winfield card is valued at $50, whereas an Al Kaline from the same set is $10. Both players achieved remarkably similar achievements, but Winfield's card is a RC while Kaline was near the end of his career. The three most valuable RCs from 1974 are Dave Winfield, Dave Parker, and Ken Griffey, all of whom are black. In 1973 the top rookies were Mike Schmidt and Goose Gossage, while in 1975 they were George Brett and Robin Yount, all of whom are white. Also note that the former is far less likely to observe randomly, because roughly two thirds of MLB players in their sample were white.

To our knowledge, the only existing work that addresses information uncertainty comes from a pair of studies by Gabriel, Johnson, and Stanton (1995, 1999; hereafter GJS95 and GJS99). They criticize previous work for ignoring how cards of active players are valued, as well as the premium associated with RCs. To account for information uncertainty they obtain card prices from sets where many of the players are still active. They analyze only the values of RCs, making their work closely related to our own. Using 1992 values of RCs from 1984-1990 Topps sets, GJS95 finds no evidence of racial discrimination. GJS99 reverses the null conclusion from GJS95, showing discrimination surfaces once performance uncertainty is removed.

Their explanation is relevant since it relates to our extension. Because they have no reason to believe racial discrimination was more prevalent in 1994 than it was in 1983, they argue:

> "Another possible explanation for the observed results centers on customer expectations. Race might enter the formation of collector expectations when a player starts his career. If collectors perceive nonwhites to be better athletes than whites, they might expect that nonwhites will have superior career statistics compared to similar white players. However, some collector's may simultaneously have negative biases against nonwhites strictly based on race. At the beginning of a player's career, card value would then capture these two effects; the empirical issue is which effect would dominate" (pg. 1334).

While this reasoning is consistent with our results and with our framing of the relevant tradeoffs facing collector's purchasing RC at their initial release point, the majority of uncertainty over player ability *has already been removed* in their 'active player' samples, where their average player has already completed six MLB seasons. While exceptions exist, most players who achieve poor, average, or exceptional outcomes over the first several years of their career continue at those respective levels. Our focus on RC prices at the point of initial release is designed to understand price formation when these critical early years of performance *have not yet been observed*.

Our study addresses the lack of attention given to the role of uncertainty in this literature by exploring the determinants of MLB issued RC values in two distinct environments – one a full information environment (post-retirement) that mirrors previous work and the other a limited information setting (at RC release). We take advantage of a unique data set that has the first reported price in the Beckett price guide for all RCs released during 1986-1993. By using the price at release, we capture the point when uncertainty over a player's career performance is greatest. If collectors are ever going to base their valuations on any factors *not related to already realized MLB achievements*, as GJS95 and GJS99 argued was likely, it should be at the point of

initial release. We also examine card prices seventeen years later to see if the effects of race change. At this point, prices have converged to a point of relative stability, as most players have retired or are close to retirement. In addition, by examining seventeen year out prices (i.e., between 2003-2010) we avoid issues concerning any lingering effects of the player's strike and subsequent owner lockout in 1994-1995, as well as any confounding effects from the documented baseball card price bubble in the early-1990s.[1]

III. Data

Most investigations prior to our own used data from the industry's early era, when Topps was the only set produced and RC identification was straightforward. The introduction of competitors and multiple product lines for each manufacturer complicated the definition of a "rookie card." There were three types of "rookie card" designations that emerged during our sample period. The standard RC definition refers to the first chronological appearance of a player on a card that was nationally available. As an example, Tim Raines first appeared in 1981 Topps and Donruss, and both are considered RCs. However, many products were not nationally available, as companies started releasing boxed sets following the completion of the regular MLB season that updated the set they released prior to the season. These sets were sold only by hobby stores that focused on sports cards. Players in these sets were typically those traded during the season or newly promoted players that had not appeared in the regular set. The Extended Rookie Card (XRC) designation noted cases where players did not have a previously made nationally available card. For example, Bo Jackson has XRC cards found in 1986 Donruss the Rookies, 1986 Fleer Update, and 1986 Topps Traded. Players with XRCs could still have the more traditional RCs from the

next year's full sets. When Bo Jackson appeared in 1987 Donruss, Fleer, and Topps, these cards carried the traditional RC designation. [2]

The final classification arose due to differences in player selection by the manufacturers. For instance, Tim Raines had 1981 Donruss and Topps cards, but not Fleer, making his 1982 Fleer release the first from that company. Hence, the Raines 1982 Fleer carries the designation FFC (First Fleer Card). Each manufacturer carried its own F*C designation, with * representing the company name. When the manufacturers began releasing multiple product lines throughout the calendar year in the 1990s the XRC and F*C distinctions became obsolete. Roughly two-thirds of cards released from 1986-1991 are denoted RCs, while 100% are RCs for 1992 and 1993.

Our 6,026 observations come from RC, XRC, and F*C cards released from 1986 to 1993.[3] The initial secondary market price of these cards is taken from the card's first listing in Beckett, which is a monthly price guide that is generally regarded as the industry standard. We use the period 1986-1993 for three reasons. First, the proliferation of sets was moderate compared to later years, such that Beckett still reported all RC prices each month. Second, in the mid-1990s, Topps began to use its Bowman brand to dominate the RC market by using its preferred licensing agreement to consistently release cards of players *before* they played in the majors. Finally, a "post career" price, mirroring the timing used in other papers on this topic, is needed for comparative purposes. Our goal was to obtain values at a point when player's careers were complete or nearing completion. This motivated using 1993 as a cutoff, such that 2010 values became the last 17 year out value pairing. We begin with 1986 because Beckett, the current and longtime industry standard for publishing card values, began in late 1984, and took some time to establish its reputation.

For post-career prices, we use the annual Beckett Almanac or annual Beckett Price Guide released 17 years following the card's initial date. The annual guides are used because Beckett stopped reporting prices in its monthly guides for many RCs to conserve space. For all but a handful of players, the 17 year mark occurs later than retirement, such that prices are highly stable.[4] Over time, the designations that remained important are the RC and XRC designations. For a very small number of cases, cards changed designation based on shifting market opinion. For example, cards from extended sets up through 1991 carried the XRC designation, but with their growing popularity, Beckett eventually replaced it with the traditional RC designation.

Because the contributions of our paper relate to the novelty, duration, and size of our database, the following section presents several stylized facts that describe it in greater detail. Between 1986 and 1993, Beckett's initial listings yield 6,026 individual rookie cards. The number of distinct players in the database is 1,910, as a player can contribute more than one observation. In addition to the initial and final prices, each set's common card price at both the initial and final listing are recorded. Common cards are the lowest valued cards in a set. Following previous studies, we use common card prices to control for overall product quality/desirability, which differs across sets. For instance, the final prices for RC from 1988 Score and 1992 Bowman have common card prices of 5 cents and 50 cents, respectively. The 1992 Bowman set had a lower print run and is general held within the hobby as being the more desirability of the two products. Measuring RC prices relative to the common card price for their set controls for these otherwise unmeasured characteristics.

In addition to RC pricing data from Beckett there are three other sources of data for the study. These sources are the physical cards themselves, player statistics from the Lahman database, and MLB draft data from www.thebaseballcube.com. The player's team, position, and race are taken

directly from the actual card. The team and position are straightforward, as we simply record the information as given on the card. Pictures on the cards, as well as biographical information on the card (e.g., birthplace) are used as a starting point to assign race as either black, Hispanic, or white.[5] Players born in Mexico, South America, Central America, and the Caribbean are classified as Hispanic. For players born in other countries, race is determined using the pictures on the cards. Note that our goal is not to provide a precise determination of racial/ethnic lineage, but to accurately reflect the perceived race of the player, based only on the visual and biographical information available on the card. This mirrors the situation faced by consumers in this market. For a very small number of cases (twelve) where race was deemed to be unclear (meaning the authors' independent assessments of the player's race differed) the players were dropped from the analysis.[6]

Player performance statistics come from the Lahman database, which has standard pitching and batting statistics for all players who have appeared in at least one Major League Baseball game, as well as biographical information. Our data contains 268 players who appear on an RC, but never played in an MLB game. For those cases, year of birth is obtained from www.baseball-reference.com. The player statistics used in our estimation are discussed in Section V.

Amateur draft data is taken from www.thebaseballcube.com. Draft position is used as an indicator of perceived player talent when few professional performance statistics are available. On rare occasions, a drafted player does not sign a contract with the team selecting him – instead reentering the next draft or holding out until the drafting team's rights expire. For our analysis, we record draft data when a player is drafted and *does sign*, recording the draft type, year, draft round, and overall pick number in the draft. For players who sign their initial contracts as amateur free agents or are purchased from teams from other countries or from minor league

teams, we record the year of their signing/purchase as the year in which they began playing professionally. For the few players who did not appear in an MLB game *and* were also not drafted, there is no official signing date reported by www.baseball-reference.com, so we simply assign the first year they played minor league baseball.

Of the 6,026 cards in the original sample, 41 are dropped due to the lack of consensus over race we described above, 1 is dropped because he was the only player of Asian heritage in the data, and 7 are dropped because the person depicted was not an MLB player (i.e., was a manager or coach). This leaves 5,977 observations over eight years. Whenever possible, our descriptive statistics, stylized facts, and regression analyses use these 5,977 cards. For 79 cards, we face the problem that they are not listed in the annual guides, from which we take final prices. As such, they are not able to enter our final price or change in price models.

IV. MLB Rookie Card Stylized Facts

To date, the most comprehensive work focusing on RCs comes from GJS95 and GJS99, who consider RCs from 156 hitters and 134 pitchers. Because our data represents a twentyfold increase over this, and because we are the first to obtain initial prices and consider the selection process for set inclusion, a detailed documentation of our data is merited.

Figure 1 provides the mean values for initial and final RC price ratios by race. The price ratios are calculated as the RC price divided by the common card price of the set at the same time. The panel on the left displays price ratios for all cards in the data, while the panel on the right displays the same information, limiting the sample to those cards which maintained their RC or XRC designation through the final pricing year.

Using the left panel to compare mean values of all cards across racial classifications, initial RC prices of black players were about 7 times common value, while white and Hispanic players fall between 4 and 5 times common value. Comparing initial and final price ratios, cards of black players saw a small increase in value, while Hispanic players saw a small decline. White players were the only group with considerable movement, as they lost roughly one third of their original value premium. Paired t-tests confirm the only statistically significant difference is white players compared to both other races.

Additionally, it is worth understanding the pricing dynamics within the subset of cards that maintained RC designation status over time. The right panel shows that cards of black and Hispanic players both increased – with slightly larger appreciation among black players. The average price ratio for black players moved from about 8.5 times more valuable than common value to over 11 times more, while the ratio for Hispanic players went from around 6 times more valuable, to just over 7. White players started with the lowest values, at about 5 times the common value, and were the only race experiencing a declining ratio over time.

Because we are the first study to obtain RC prices at their initial release, as well as over time, one objective is to document the differences in price trends across racial categories. Of course, it is natural to wonder why these differences are present. Recall that GJS95 argue collectors may have higher expectations for young minority players if minority players from earlier periods became superstars at higher rates than whites. Evidence based on selection to All-Star games in the 1950s-1970s, and career performance measures (both our own and data used in previous studies), suggests this pattern is present in MLB. Nardinelli and Simon (1990) used a binary variable called STAR, which tracked whether players were worth above-common values or not. Mean values in their sample (which contains players active in the 1960s and 1970s) were 0.49,

0.23, and 0.27 for black, Hispanic, and white players, respectively. Before we show more precisely how these patterns surface in our data, we stress that interpreting this difference is complicated by previous studies suggesting black players were discriminated against while trying to attain MLB status. If black players face a harder-than-average path to the majors, the players who make it past these higher barriers will, in turn, be better-than-average. Another challenging complication is that collectors of baseball cards may or may not *simultaneously* have racial biases regarding willingness to pay that may operate alongside their potential desire to take advantage of meaningful statistical discrimination patterns.

We are interested in the possibility that differences in RC price trends over time could be sensitive to whether or not MLB card companies have different thresholds for releasing cards of younger black, Hispanic, and white MLB players. If present, selection biases of this type would exacerbate any differential experiences affecting the path to making the majors. For this reason, we take up the issue of player selection for set inclusion.

Player Selection for Set Inclusion.—If card manufacturers utilize differential selection criteria for including young black, Hispanic, and white players in their products, it would complicate how to best interpret our stylized facts regarding price trends over time. One possibility is that manufacturers are selecting more prospects (i.e., minor league players without MLB experience) than players who have already appeared in MLB, and that this choice interacts with any effect race may have on the chances of being promoted from the minors to the major leagues.

To examine this possibility, we record the number of players that were selected for a card set, but who were never promoted to the major leagues. We also record how many players were included in a set without having prior major league experience (i.e., whether they did or did not eventually reach MLB). White players comprise 68 percent of the total sample of cards, black

players 17 percent, and Hispanic players 15 percent. Moving from all cards in our sample to all players in our sample we see no discernible differences for players by race. From this evidence, we conclude players were selected for set inclusion at similar rates regardless of race. Regarding players who never appeared in an MLB game, the percentage of white players is consistent with their overall representation (between 67 and 69 percent in all categories). However, percentages for black players increase slightly (4 to 7 percentage points) and the compositional shift is offset by a decrease in Hispanic players. Thus, it appears black players are slightly more likely to have a RC without actually appearing in a major league game. This initial piece of evidence suggests that, conditional on other discriminatory behaviors that may or may not be occurring, card producers do not seem to have systematically higher thresholds for inclusion of black players.

It is also worth noting how many players without prior MLB experience were included in sets, regardless of whether they would eventually play in MLB (as most did). Figure 2 shows, by year and race, the percentage of players who debuted in MLB before their RC was released. The panel on the left treats each card as an observation, whereas players are considered in the right panel. Through 1988, about 95 percent of RCs show a player with prior MLB experience. In 1989, the fraction of players with experience at release drops to historical lows for all three racial categories. Each then *again* hits new unprecedented lows in 1990, 1991, 1992, and 1993, respectively. Thus, manufacturers clearly began to release more cards of players without experience over the last half of our data.

Focusing on potential differences across racial classifications, black players had the steepest decline in terms of experience, with only 27 percent of cards and 15 percent of players having MLB experience prior to the release of their RCs from 1993. While Hispanic and white players did not see as dramatic a decline, they still fell to around 60 percent of their cards and 33 percent

of the players having MLB experience. The smaller decline in these numbers for Hispanic players may be due to contractual issues. In the latter years of our data, members of the US Olympic baseball team and those who were first round draft picks in baseball's summer draft were commonly featured on cards. As many Hispanic players spend their teen years playing outside the US, this limits the ability of companies to include them in these subsets. On the whole, the evidence seems to indicate each racial group experienced roughly similar thresholds for set inclusion, and that these thresholds changed in similar patterns over time for each group.

Player Statistics.—Another possibility is that MLB players have varied achievements by race. If so, these differences may be evident for young players. We find the answer to this question depends on two factors: whether player quality is measured as productivity per season or by total career achievements (i.e., respecting longevity), and whether one focuses exclusively on measures of central tendency or on the tails of the performance distribution.

Regarding the former, hitter's average statistics per 162 games show little difference across racial groups, except for base stealing, where black hitters have higher averages. However, measures of longevity reveal significant differences. Black and Hispanic hitters play roughly 16 percent and 25 percent more career games than white hitters.

Minor differences across racial groups occur for pitchers, although these differences are primarily caused by the pitching role. Black pitchers have significantly shorter careers when measured by innings pitched, but more career games, reflecting that they may be used more as relief pitchers than starting pitchers. Other than these differences, pitchers of each racial classification have similar statistics per 200 innings pitched.[7]

To examine whether the tails of the distribution could influence collectors expectations, we separate hitters and pitchers into various percentiles based on achievements, comparing the racial

composition of these subgroups with the overall sample. Hitters are placed into percentiles based on career games played. To account for the disparate usage patterns of starters and relievers, pitchers are ranked based on the product of career innings pitched and games played.

Figure 3 shows the percentage of each quartile by race for hitters and pitchers separately. We find a U-shaped pattern for black hitters, who are about 30 percent of the top and bottom quartiles, but 22 and 26 percent of the middle quartiles. White hitters are half of the top quartile, compared to 55-60 percent for each of the other three. Hispanic hitters are consistently 15 percent of the first, second, and third quartiles, increasing to 20 percent of the top quartile. The fraction of white pitchers is stable across quartiles, always between 80 and 83 percent. There is a decline in the percentage of black pitchers across quartiles, which is offset by a similar increase in Hispanic pitchers. On the whole, Figure 3 indicates that racial composition by performance is highly nuanced. Relative to the overall sample, white hitters are less likely to reach the top quartile, whereas black and Hispanic players are more likely. For pitchers, only minor differences surface. Finally, because black pitchers are found to be more likely to be relief pitchers, it is difficult to draw firm conclusions in this case.

We pursue a more refined view of the top quartile because previous work has stressed the idea that star performing players command a considerable premium. We also use this view to address whether or not investors and card companies are correctly predicting which players will have better careers at the time of RC release, and to examine the relationship between eventual player career accomplishments and final RC values. Table 2 shows, for hitters and pitchers respectively, the number of players, number of RCs per player, mean initial price ratio, and the mean final price ratio for the overall sample and by racial group, for various percentiles. The first three quartiles remain as presented in Figure 3, but the top quartile is split into the 75^{th}-90^{th} percentiles,

the 90th-95th percentiles, and the 95th-100th percentiles. If investor's expectations regarding the chance of attaining eventual superstar status carry the potential to be affected by race, we should see deviations from the full sample surface in the upper tail.

For the full hitter sample, the number of cards per player and mean initial price ratios increase monotonically as percentiles considered rise, with final price ratios displaying nearly the same. The potential benefit from statistical discrimination is seen through a comparison of the mean initial and final price ratios. As expected, the value distribution is far less severely skewed in the environment where uncertainty prevails. While players achieving below average outcomes have final RC values that barely exceed common card prices, their initial values are nearly three times as high. Moreover, highly successful players reaching the 90th-95th percentiles experience no increase in value over time. The payoff from purchasing RC of players who will eventually attain star status is seen in the 95th-100th percentiles, where value ratios move from about 14 at initial release to nearly 34 at final prices. Hitters in the bottom quartile have roughly 1.5 cards issued, while hitters in the top quartile average 4.5. Manufacturers, by printing more cards of the players who ultimately achieve the most, as well as collectors, by paying higher initial prices for those players, seem to be aware of which players will eventually perform better.

Conditional on race, these price differentials are more severe for black players, slightly less severe for Hispanic players, and significantly less severe for white players. RCs of the 21 white hitters who reach or exceed the 95th percentile experience a slight increase in value, moving from 15.3 to 18.8 times more valuable than commons. RCs of the 13 black players meeting the same qualification move from 16.2 to 71.5 times as valuable, while the ratio for the 14 Hispanic players attaining the top category was 10.1 to 23.5. Regarding race and potential statistical discrimination, minorities constitute 45 percent (438 out of 976) of the overall pool of hitters, but

represent 56 percent (27 out of 48) of the players attaining the 95th-100th percentiles. Interestingly, the shift is slightly stronger for Hispanic players than for black players.

To account for the difference in starting and relief pitchers, all pitchers are placed in their percentile group by the product of their innings pitched and career games.[8] Taking pitchers as a whole, the market succeeds in supplying more RCs of pitchers who eventually achieve more, and in pricing those RCs at initially higher values. However, while the market is correct on average, there is less accuracy than what we observe for hitters. The mean initial value ratio of a pitcher destined for the bottom quartile (3.42) is indistinguishable from the 50th-75th percentile (3.47), and only 50 percent less than the 95th-100th percentiles. Compare this to hitters, where the 50th-75th percentiles have a 50 percent premium over the bottom quartile and the 95th-100th percentiles are nearly 400 percent higher. This difference confirms what baseball scouts have long claimed – it is harder to forecast top pitching talent than it is to estimate who will be the most productive hitters. Also, the final price ratios show that, once performance uncertainty is gone, the value premium for top quality players resurfaces – although to a lesser degree than for hitters.

Regarding racial compositional shifts along the distribution, because white players comprise 82 percent of the overall pitcher sample (747 out of 912), there are few observations in the upper percentiles representing black and Hispanic pitchers. Hispanic pitchers constitute 11 percent of the overall pitcher sample (102 out of 912) and 11 percent of the top decile (10 out of 91). Black pitchers are 7 percent of the overall sample (63 out of 912), but only 4 percent of the top decile (4 out of 91). Collectively, our data contains no evidence that collectors should view the race of young MLB pitching prospects as influencing the likelihood of achieving star status. However, this does not rule out the possibility that statistical discrimination may affect pitchers as well.

V. Methodology: Contrasting Pricing Models

Recall that prices are observed at the point of initial release, and then again seventeen years later. As argued above, these environments represent markets where the information available to collectors/investors differs dramatically. In particular, there is less information available about the players upon initial release. Some players may have limited major league experience, some may have only minor league experience, and others may have no professional experience. By the 17 year out mark, most players have concluded their careers, and card buyers have far more information concerning the player's performance quality. In turn, our models that explain card values at the 17 year mark include more explanatory variables than the initial release models, and are expected to explain a much greater portion of the observed variation in card prices. Additionally, we want to investigate the factors influencing price movements over time, so we build a third model that combines price information from both points of observation.

Dependent Variables.—The dependent variable in the initial and final price models is the natural log of the ratio of the RC price to the common card price from that set.[9] This ratio helps control for set specific traits that are unrelated to player quality. Those familiar with the hobby agree that cards from some sets are more desirable because manufacturers use higher quality materials or have smaller print runs. Fortunately, these unobservable factors influence RC and common card values similarly, leaving the ratio between the two largely invariant. In the model explaining price changes over time, the dependent variable is the natural log of the ratio of the final year price ratio to the initial year price ratio. As such, if the common price has not changed, the measure reflects the ratio of the final year price to the initial year price.

Independent Variables.—We begin with a discussion of the independent variables used in our first estimations - initial price models that use all players. While most of our regressions follow

previous work by including only hitters or pitchers, we use this model as a point of comparison for others that follow, as well as to demonstrate that our results regarding player race are not sensitive to this distinction. Player specific binary variables indicate whether a player is black or Hispanic, whether the card is initially listed as an RC or XRC, whether the player is a hitter, whether the player was drafted through the MLB draft, whether the player has MLB experience at the time of the card release, and the specific position(s) listed on the card.[10] Player specific continuous variables are the age of the player at the time of card release, how long it has been since the player was signed at the time of card release, and how many cards the player contributes to our data set. Finally, a series of team level binary variables indicating the professional franchise the player is identified with on the front of the card is included.[11] Formally then, our initial price models start as a baseline GLS random effects regression:

$$(1) \qquad \ln\left(P_i/P_{C_i}\right) = \beta_0 + \beta_1 Black_i + \beta_2 Hispanic_i + \beta_3 RC_i + \beta_4 XRC_i + \beta_5 Age_i + \beta_6 Drafted_i + \beta_7 TimeSinceSigned_i + \beta_8 Hitter_i + \beta_9 NumCards_i + \beta_{10} Position_i + \beta_{11} MLBExp_i + \delta Team_i + \varepsilon_i$$

Once we move to the more traditional approach of grouping hitters and pitchers into separate regressions, the position variables are modified accordingly and the hitter variable is dropped. In estimations using only the subset of drafted players, we replace the drafted binary variables with a set of variables indicating the draft round and overall selection position in the draft.

While most of our independent variables are directly obtained from the data sources described in Section III, it is worth noting one exception. We classify a player as having MLB experience if his major league debut is at least three months prior to the initial Beckett listing of the card. For the vast majority of players this is a trivial issue – they either had a card initially listed well before their debut dates (e.g., cards of draft picks or Team USA cards) or only had cards printed

well after their MLB debut date (e.g., players like Tom Glavine, who played during the 1987 season and then had multiple RCs in 1988 products). However, once manufacturers began releasing products in the summer, it became more difficult to determine whether or not the player had MLB experience in cases where the player's card release and MLB debut were in the same season. We use a three month buffer to account for the actual timing of information in this market. For example, if a player debuted in May of 1989, and his card was initially priced in Beckett in June of 1989, he *would not* be classified as having prior MLB experience (MLB experience would be assumed only if that RC card was first listed in August 1989 or later). The reason we use a three month buffer is due to the difference between newsstand timing of the Beckett guides and when available information about the player reaching MLB would be available. Using two versus three months only changes the experience variable slightly, as only 38 observations (0.6 percent of our sample) are affected by the choice.

Our final price regressions follow equation (1) with two important extensions. First, we expand the set of independent variables based on the information that becomes available over players' careers. Because hitters and pitchers attain different accomplishments, all final price regressions are estimated separately for hitters and pitchers. Like the initial price models that separate hitters from pitchers, the binary variable indicating whether a player was primarily a hitter was dropped. In the hitter models we include career hits, doubles, triples, home runs, walks, stolen bases, and at-bats. In the models for pitchers we include wins, games, games started, complete games, shutouts, saves, strikeouts, and innings pitched. For hitters the number of games played in the postseason (division series, league championship series, and World Series) is also included, while the number of games pitched in the postseason is included for pitchers. Additionally, because we observe the same type of left censoring issue in our price data

that Nardinelli and Simon (1990) found, we complement our random effects GLS models with random effects Tobit models.[12] Finally, for consistency, the regressions modeling the change in price over the 17 year period use the same set of explanatory variables as the final price models.

VI. Results

We separately present the results of our initial price, final price, and change in price models, then offer a discussion of how the three sets of results can be collectively understood.[13]

Initial Pricing Models—Table 3 presents the results of the estimated models for all players, for hitters and pitchers separately, and only for hitters and pitchers who signed a contract as a draft pick.[14] All three estimations reveal a significant price premium on the cards of black players, as they are priced 15 percent to 17 percent higher than cards of white and Hispanic players. Cards originally listed as an RC or XRC had a 45 percent premium over cards which were listed as F*Cs for each of the manufacturers. Age at the time of release and the length of time that had passed since singing a contract had significant negative effects. Investors and collectors seem to believe young players receiving enough attention to have a card made are, on average, going to have more productive careers than players who see their first card at older ages. Compared to those with no prior MLB experience, players with MLB experience had a significant premium of about 20 percent. Taken together, the biggest premium is for young players who have played in MLB. For a moment ignoring offensive position, we find a 16 percent average premium for hitters as compared to pitchers. However, while this premium applies to offensive positions like 2nd and 3rd base, catcher, and shortstops, the premium grows (approximately doubling) at traditional 'power hitter' positions like first base, outfield, and designated hitter. Each additional card that the player appeared on in our data set (whether it was

an RC, XRC, or F*C) had a positive impact (about 8.5 percent) on initial pricing. This is consistent with our stylized facts that players of eventual higher quality (i.e., who had longer careers) had more cards released by the manufacturers, and provides suggestive evidence that seeing a player with more RC released serves as a positive signal to collectors.

In the model with all players there is no significant difference between the prices of amateur draft signees and amateur free agent signees. However, when focusing on only the drafted players, there are significant effects for those who signed in the June Regular draft, though not for those who signed in the January drafts or June Secondary draft. For those who signed in the June Regular draft, there is a positive effect for draft round, while overall draft pick number has a somewhat offsetting negative effect.

Most of our coefficient estimates have directional effects, magnitudes, and significance levels that are very similar to the results from our hitter models, and we see premiums for cards of black pitchers ranging from 12.5 percent to 17 percent. As such, we have evidence from both groups of players indicating there is a positive and statistically significant premium placed on the RCs of black players at the time of initial pricing. This finding contrasts with the literature on race and baseball cards, as coefficient estimates for binary variables indicating black players are typically negative or insignificant. Again though, our work is the first to explain RC values using only information that was available to collectors at the time of initial release, making direct comparisons with previous studies potentially misleading.

Final Price Models—Table 4 shows the results for the estimated final price models for hitters and pitchers. As hitters and pitchers accumulate different sets of accomplishments over their playing careers, we estimate all models using either hitters only or pitchers only. While we include all the performance measures in these regressions, presented results suppress those which

did not achieve statistical significance. Also, because players who never appeared in an MLB game have no performance measures, they are excluded. All models include dummy variables for a player's team as listed on the card, and models for hitters include positions as listed on the card. Recall that both random effects GLS and random effects Tobit models are estimated, due to left censoring issues with the final price ratios. To these we add two restricted information estimations, one a random effects GLS model that only includes the race variables and a final model that includes only variables present in our first year models.

Unlike our initial RC price models, the results in the final price models for both hitters and pitchers reveal no significant differences in the prices of cards for players of different races, regardless of whether or not we control for player performance. One of the reasons we find this result to be convincing is that *the same is not true* for most other variables common to both sets of regressions. Aside from the race variables, most of the other variables have similar signs, point estimates, and significance levels. One other difference is the variable for prior MLB experience, which is negative in the final price models but was positive in the initial price models. This reversal is likely driven by players who had draft pick cards issued prior to MLB experience, and then had outstanding careers, thus removing any uncertainty. Derek Jeter, Frank Thomas, Manny Ramirez, and Chipper Jones are a few notable players who fit this category.

For hitters the statistically significant player statistics are hits, home runs, stolen bases, at bats, and post season games. Hits, home runs, stolen bases, and postseason games have positive effects, while at bats have a negative effect. Taking hits and at bats together, this suggests that a higher career batting average leads to more highly valued cards. For pitchers, the only significant performance variables in the final price model are wins, games, shutouts, saves, strikeouts, and number of postseason pitching appearances. All have positive signs except for games and

shutouts. The negative effect of games, controlling for factors like wins and strikeouts, simply indicates lower quality pitchers take longer to accumulate a given level of positive achievement – in much the same way the hitter model penalized additional at bats. As for the negative effect of shutouts, it should be noted that the number of shutouts is a very small number compared to other variables for most pitchers in our data set, and that the negative coefficient on shutouts might be penalizing starting pitchers relative to relief pitchers. In general, the results on player performance are largely consistent with earlier research.

Change in Price Models—Table 5 presents the results of our price change models, where the dependent variable is the difference in log final price ratio and log initial price ratio.[15] A difference between these and our prior results is that now the race variable for black players is negative and statistically significant at conventional levels in the model for hitters. Controlling for a host of other factors including career performance, the prices of RCs for black players declined at a significantly greater rate than those of white players. On the one hand, this is consistent with our other results, because the initial price models revealed a premium placed on cards of black players, whereas the final price model did not. On the other, the result may seem puzzling, given the fact that RCs for black players *appreciated more dramatically* than cards of white or Hispanic players. Put another way, while our findings show investors initially paid *more* for cards of black players, the value of those same cards did not hold up as well over time, once we control for a constant level of achievements moving forward.

One reasonable explanation supporting these seemingly contradictory results is the idea that collectors may raise their level of expectations regarding a player's future performance when the player is black – illustrating the manner in which our results suggest investors seem to form their decisions in uncertain environments in ways that account for statistical discrimination patterns.

Of course, these expectations line up with realized outcomes in the data. Recall that, relative to their rates of prevalence in the overall sample, black hitters were overrepresented at the highest and lowest quartiles of player achievements. The overrepresentation in the lowest quartile should be of little concern to investors if the final RC prices of those with poor and average careers end up very close to one another (as we observe they are). However, the overrepresentation in the top quartile is more meaningful – as black hitters carry higher than average probabilities of eventually becoming stars. Put another way, if the performance of black hitters carries a larger variance, this is meaningful, given the way the market asymmetrically values player's achievements, (i.e., placing a large premium on superstars relative to even high quality players, and grouping a majority of the eventual value distribution together).

Interestingly, this collective pattern of results (i.e., positive and significant in the initial price models, insignificant in the final price models, with negative and significant in the price change models) is also seen for the position dummies associated with playing first base, outfield, and designated hitter. As such, it appears race is one, but not the only, observable characteristic that gives investors an opportunity to take advantage of meaningful statistical discrimination patterns. In fact, our story suggests any observable factor that correlated with the long run probability of achieving superstar status (and in turn, the highest final RC values), should *temporarily* induce a higher willingness to pay in settings affected by information uncertainty that *does not later translate* into a residual willingness to pay for the presence of the trait itself. This is why the most surprising aspect of our collective findings is that we do not pick up the same pattern for Hispanic hitters, who meet these criteria, whereas we see it for black pitchers, who do not. For Hispanic hitters, we find the same direction of effects, although they are of smaller magnitudes and do not achieve statistical significance. Regarding black pitchers, our results are more

difficult to explain, indicating collectors may incorrectly associate likely distributional outcomes of hitters and pitchers. Finally, all remaining covariates have signs similar to those in the final price models, although the player's time since signing is now insignificant.

VII. Discussion

The seminal work of Nardinelli and Simon (1990) took up the controversial issue of customer based racial discrimination by investigating the impact of race on baseball card values, finding significant premiums for cards of white players relative to black and Hispanic players. Most of the work that followed either confirmed this finding or failed to uncover evidence of premiums related to race. Using a novel data set that contains initial and final prices for a large sample of rookie cards issued from 1986 through 1993, we show the effects of race on RC values are nuanced. Upon initial release, when players have little/no MLB experience and are unproven commodities, we find cards of black players carry a 15 to 19 percent premium over comparable white players. When analyzing prices of the same cards seventeen years later, and controlling for eventual career achievements, we find no significant price effects based on race. We connect these findings by estimating models that consider the change in price over time, where we find a significant penalty accrues for black player's RCs. This implies that when RC of white and black players start with equal values, black players must attain a significantly higher level of career achievement to retain that price, relative to white players.

Given previous results, a positive premium surfacing for black players at the time of initial pricing is meaningful. After exposing this basic result to a number of different modeling choices and sensitivity tests, we explain how it is consistent with two documented stylized facts. The first is that the relationship between final RC values and player performance is highly asymmetric.

Final values for average players are much closer to those of the worst players than they are to values of the best players. The second is that black hitters have a significantly higher variance regarding their eventual career achievements than whites or Hispanics. Black hitters are more likely to fall in the highest and lowest portions of the achievement distribution than they are to be located near the middle. We argue forward-thinking investors are likely to form expectations using statistical discrimination in settings where information uncertainty is present, but not in settings where the uncertainty has been removed. Using the same line of reasoning, two puzzles are uncovered by our work. The first is that the same pattern (i.e., an initially significant premium that disappears by career completion) is seen for black pitchers, even though our player career achievement data does not display an overrepresentation of black pitchers in the tails of the performance distribution. The second, and perhaps more meaningful, is that the same pattern is not observed for Hispanic hitters, where the data reveal a similar overrepresentation in the right tail, meaning the same investment technique would be beneficial.

Importantly, because none of our final price models show significant lingering effects of race, we do not find evidence that suggests purely discriminatory behavior influences the baseball card market. However, we do find the race of unproven MLB prospects correlates with other meaningful differences in future performance outcomes, and that investors seem to form their willingness to pay in uncertain environments, in ways that account for statistical discrimination. This pair of results is consistent with prior studies that suggest racial discrimination still plays an important role in MLB labor contracts.

REFERENCES

Andersen, Torben, and Sumner J. La Croix. 1991. "Customer Racial Discrimination in Major League Baseball." *Economic Inquiry* 2:665-677.

Ayres, Ian, Mahzarin Banaji, and Christine Jolls. 2011. "Race Effects on eBay." Yale Law School Working Paper.

The Baseball Cube. 2012. "2012 Draft Register." http://www.thebaseballcube.com/data/draft_register.asp

Becker, Gary. 1971. The Economics of Discrimination. 2nd edition. University of Chicago Press, Chicago: USA.

Bellemore, Fred A. 2001. "Racial and Ethnic Employment Discrimination: Promotion in Major League Baseball." *Journal of Sports Economics* 2: 356-368.

Bergmann, Barbara R. 1974. "Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex." *Eastern Economic Journal* 1: 103-110.

Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94: 991-1013.

Burnett, Nancy J. and Lee VanScyoc. 2004. "Discriminating Buyers of Baseball Cards: Does Race Affect Value?" *Journal of Economics and Economics Education Research* 3: 101-110.

Burnett, Nancy J. and Lee VanScyoc. 2012. "Changing Levels of Discrimination in the Market for Baseball Cards." *Journal of Economics and Economics Education Research* 13: 101-110.

Butler, Richard J. 1982. "Estimating Wage Discrimination in the Labor Market." *Journal of Human Resources* 17: 606-621.

Curtis, James E. and John W. Loy. 1978. "Positional Segregation in Professional Baseball: Replications, Trend Data, and Critical Observation." *International Review for the Sociology of Sport* 13: 5-23.

Depken, Craig A. and Jon M. Ford. 2006. Customer-based Discrimination against Major League Baseball Players: Additional Evidence from All-Star Ballots." *Journal of Socio-Economics* 35: 1061-1077.

Findlay, David W. and John M. Santos. 2012. "Race, Ethnicity, and Baseball Card Prices: A Replication, Correction, and Extension of Hewitt, Munoz, Oliver, and Regoli." *Econ Journal Watch* 9: 122-140.

Fort, Rodney and Andrew Gill. 2000. “Race and Ethnicity Assessment in Baseball Card Markets.” *Journal of Sports Economics* 1: 21-38.

Gabriel, Paul E., Curtis D. Johnson, and Timothy J. Stanton. 1995. “An Examination of Customer Racial Discrimination for Baseball Memorabilia.” *Journal of Business* 68:215-230.

Gabriel, Paul E., Curtis D. Johnson, and Timothy J. Stanton. 1999. “Customer Racial Discrimination for Baseball Memorabilia.” *Applied Economics* 31:1331-1335.

Gwartney, James and Charles Haworth. 1974. “Employer Costs and Discrimination: The Case of Baseball.” *Journal of Political Economy* 82: 873-881.

Hanssen, F. Andrew and Torben Andersen. 1999. “Has Discrimination Lessened Over Time? A Test Using Baseball’s All-Star Vote.” *Economic Inquiry* 37: 326-352.

Hewitt, John D., Robert Munoz Jr., William L. Oliver, and Robert M. Regoli. 2005. “Race, Performance, and Baseball Card Values.” *Journal of Sport and Social Issues* 29: 411-425.

Holmes, Paul. 2011. “New Evidence of Salary Discrimination in Major League Baseball.” *Labour Economics* 18: 320-331.

Holzer, Harry J. and Keith R. Ihlandfeldt. 1998. “Customer Discrimination and Employment Outcomes for Minority Workers.” *Quarterly Journal of Economics* 113: 835-867.

Holzer, Harry J., Steven Raphael, and Michael Stoll. 2006. “Percieved Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers.” *Journal of Law and Economics* 49: 451-480.

Lahman, Sean. 2011. “Lahman’s Baseball Database.” http://www.seanlahman.com/baseball-archive/statistics/

List, John A. 2004. “The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field.” *Quarterly Journal of Economics* 119: 49-89.

McGarrity, Joseph, Harvey D. Palmer, and Marc Poitras. 1999. “Consumer Racial Discrimination: A Reassessment of the Market for Baseball Cards.” *Journal of Labor Research* 20: 247-258.

Nardinelli, Clark, and Curtis Simon. 1990. “Customer Racial Discrimination in the Market for Memorabilia: The Case of Baseball.” *Quarterly Journal of Economics* 105: 575-95.

Neumark, David. 1999. “Wage Differentials by Race and Sex: The Roles of Taste Discrimination and Labor Market Information.” *Industrial Relations* 38: 414-445.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." *American Economic Review* 101: 1410-1435.

Pascal, A.H. and L.A. Rapping, 1972, "The Economics of Discrimination in Organized Baseball", in Pascal A.H. (ed.) *Racial Discrimination in Economic Life* (Lexington, Mass: Lexington Books) 119-156.

Primm, Eric, Nicole L. Piquero, Robert M. Regoli, and Alex R. Piquero.2010. "The Role of Race in Football Card Prices."*Social Science Quarterly* 91:129-142.

Primm, Eric, Nicole L. Piquero, Alex R. Piquero, and Robert M. Regoli. 2011. "Investigating Customer Racial Discrimination in the Secondary Baseball Card Market", *Sociological Inquiry* 81: 110-132.

Regoli, Bob. 1991. "Racism in Baseball Card Collecting: Fact or Fiction?" *Human Relations* 44: 255-264.

Regoli, Robert M., Eric Primm, and John D. Hewitt. 2007. "Men and Boys and the Price of their Toys: Race and the Value Football Cards." *Social Science Journal* 44: 563-571.

Scahill, Edward M. 2005. "A Reinvestigation of Racial Discrimination and Baseball Cards." *Eastern Economic Journal* 31: 537-550.

Scully, Gerald W. 1973. "Economic Discrimination in Professional Sports." *Law and Contemporary Problems* 38: 67-84.

Stone, Eric W. and Ronald S. Warren. 1999. "Customer Discrimination in Professional Basketball: Evidence from the Trading Card Market." *Applied Economics* 31: 679-685.

Yinger, John. 1986. "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act." *American Economic Review* 76: 881-893.

Zillante, Arthur. 2008. "A Review of the Post-WWII Baseball Card Industry," in the *Business of Sports (Volume 3)*, eds. Brad Humphreys and Dennis Howard, Praeger Publications: Westport, CT, 241-272.

Zillante, Arthur. 2012. "Bubblegum Card Bubbles." UNC Charlotte working paper.

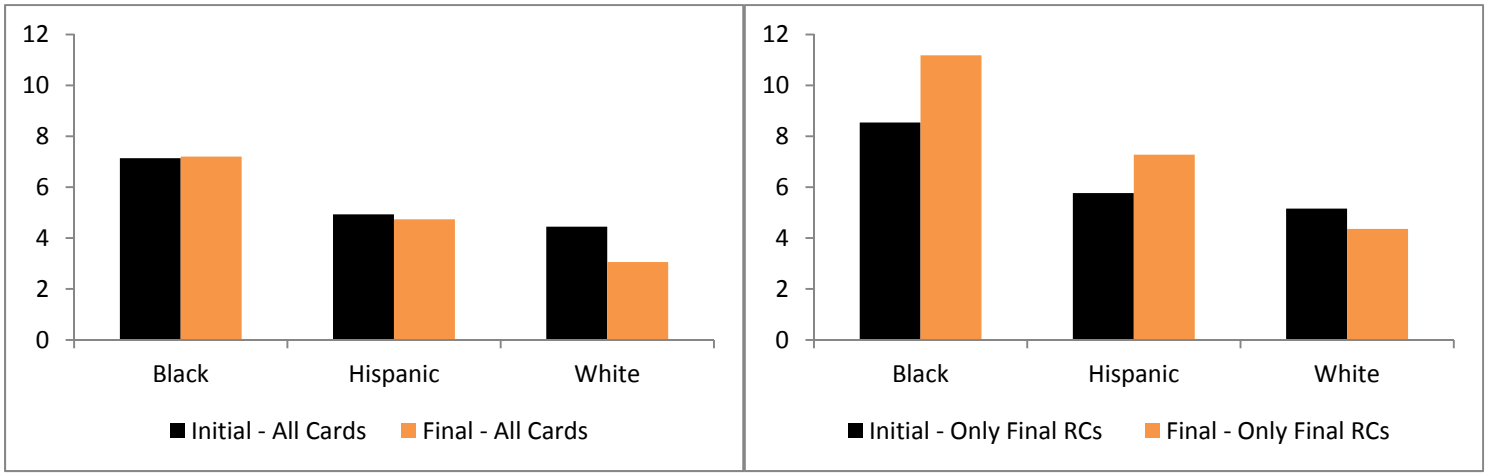

Figure 1—Initial and final price ratios for all cards in the sample (left) and cards that retaining RC or XRC status in the final time period for prices (right) by race

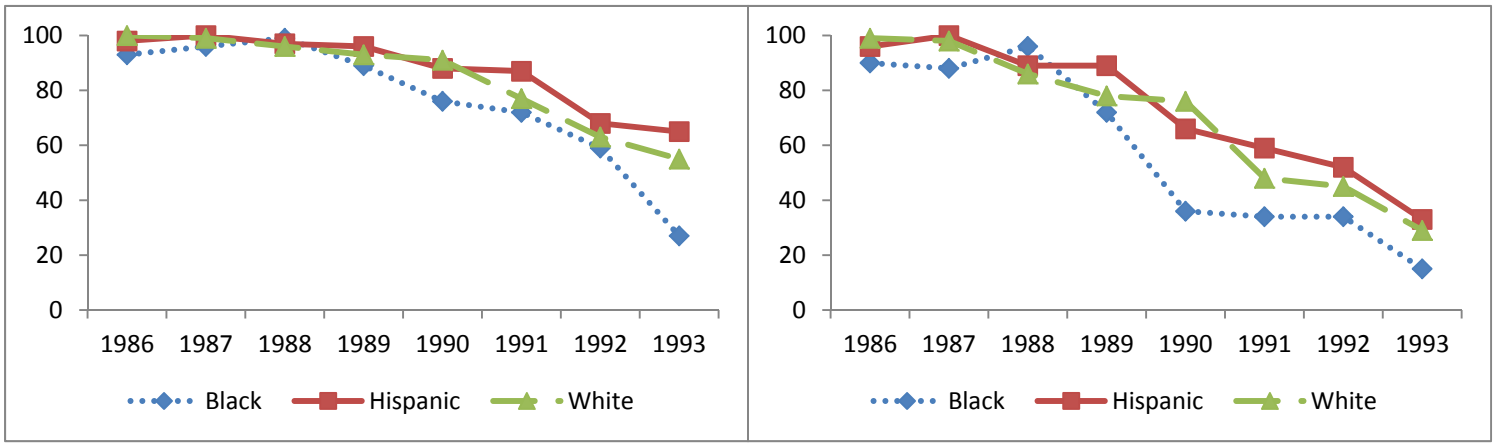

Figure 2—Percentage of cards (left) and players (right), by race and year, with MLB experience prior to RC release

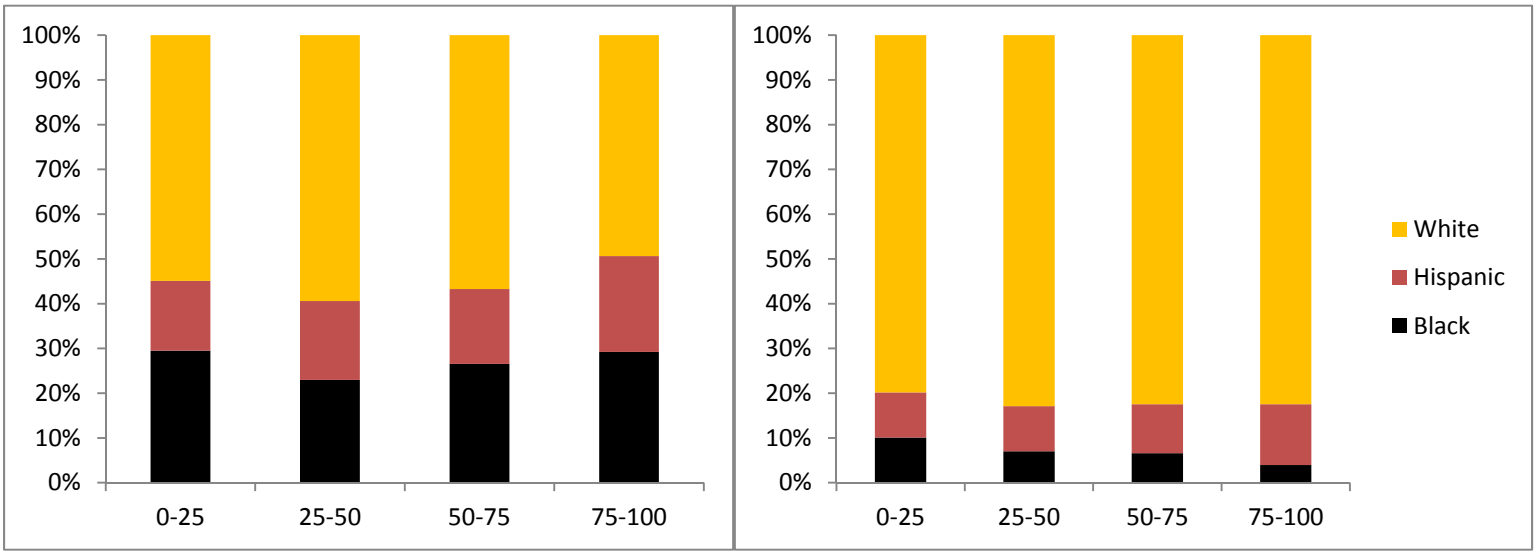

Figure 3 – Percentage of players, by race, for each performance quartile. Hitters are in the left hand panel and pitchers are in the right hand panel.

Table 1 – Summary of prior investigations of the effects of race on baseball card values.

Paper	Card Years	Sample	Beckett	Found Discrimination Against
Nardinelli & Simon (1990)	1970	567	1989	Hispanic hitters; black pitchers
Andersen & LaCroix (1991)	1960-1961	534	1982	average and below average black pitchers
	1977	584	1985	black hitters
Regoli (1991)	Various (RCs)	25	1990	none
Gabriel, et al. (1995)	1984-1990 (RCs)	290	1992	none
Gabriel, et al. (1999)	1974-1992 (RCs)	140	1983	none
	1974-1992 (RCs)	140	1994	black and Hispanic hitters
McGarrity, et al. (1999)	1974	522	1994	none
Fort & Gill (2000)	1987	457	1987	black and Hispanic hitters; black pitchers
Burnett & VanScyoc (2004)	1960-1969	2833	1992	black and Hispanic hitters
Hewitt, et al. (2005)	Various (RCs)	51	2003	none
Scahill (2005)	1970	578	Various	black and Hispanic hitters (late 1980s), none (1990s)
Burnett & VanScyoc (2012)	1960-1969	2770	1981	none
	1960-1969	2770	2008	black hitters
	1986	374	2008	none
Findlay & Santos (2012)	Various (RCs)	51	2003	none
	Various (RCs)	51	2010	none

Table 2 – Number of players, cards per player, and mean initial and final price ratios by race and across various percentiles. Data for hitters is on the left while pitcher data for the similar classification is on the right.

Percentile	# Players	Cards per player	Mean initial price ratio	Mean final price ratio	# Players	Cards per player	Mean initial price ratio	Mean final price ratio
		All Hitters				All Pitchers		
0-25th	244	1.54	3.56	1.49	228	1.61	3.42	1.40
25th – 50th	244	2.88	3.67	1.24	228	2.68	2.86	1.15
50th – 75th	245	4.03	5.77	1.91	228	3.89	3.47	1.36
75th – 90th	145	4.25	5.95	3.95	137	4.16	4.83	2.54
90th - 95th	50	4.46	9.92	9.34	46	3.96	4.18	2.96
95th - 100th	48	5.69	13.92	33.90	45	4.09	5.11	13.73
		Black Hitters				Black Pitchers		
0-25th	72	1.61	4.34	1.47	23	1.61	7.24	1.57
25th – 50th	56	2.86	3.95	1.25	16	2.69	2.87	1.07
50th – 75th	65	4.26	8.91	2.45	15	3.13	4.71	1.37
75th – 90th	43	3.70	6.54	3.89	5	5.40	3.89	2.17
90th - 95th	15	5.27	9.71	5.16	1	5.00	3.15	3.60
95th - 100th	13	5.46	16.16	71.49	3	5.67	7.37	5.00
		Hispanic Hitters				Hispanic Pitchers		
0-25th	38	1.53	3.42	1.36	23	1.78	3.14	1.12
25th – 50th	43	3.14	3.88	1.14	23	2.70	2.65	1.09
50th – 75th	41	3.93	4.56	1.39	25	3.28	3.97	1.52
75th – 90th	25	3.36	5.56	6.78	21	3.95	4.07	1.76
90th - 95th	13	4.08	7.91	8.02	5	4.20	5.19	2.55
95th - 100th	14	6.00	10.09	23.48	5	1.40	5.69	33.38
		White Hitters				White Pitchers		
0-25th	134	1.50	3.15	1.54	182	1.58	2.97	1.42
25th – 50th	145	2.81	3.48	1.27	189	2.67	2.89	1.16
50th – 75th	139	3.96	4.54	1.78	188	4.04	3.34	1.34
75th – 90th	77	4.84	5.79	3.32	111	4.14	5.02	2.70
90th - 95th	22	4.14	11.27	13.98	40	3.90	4.08	2.99
95th - 100th	21	5.62	15.30	18.77	37	4.32	4.84	13.80

Table 3 – Regression results modeling initial prices. Model (1) provides estimates for all players, both hitters and pitchers. Models 2-4 provide estimates for only hitters, only drafted hitters, and only hitters using just race variables. Models 5-7 are identical to models 2-4, except for pitchers. Standard errors are in parentheses and ***, **, and * represent significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Black	0.141***	0.149***	0.158***	0.203	0.119**	0.159**	0.107
	(0.034)	(0.043)	(0.047)	(0.046)	(0.060)	(0.065)	(0.062)
Hispanic	0.027	0.052	0.070	0.056	-0.009	0.009	0.001
	(0.043)	(0.062)	(0.082)	(0.054)	(0.059)	(0.087)	(0.049)
RC Initial	0.370***	0.391***	0.393***		0.348***	0.353***	
	(0.016)	(0.024)	(0.027)		(0.020)	(0.022)	
XRC Initial	0.333***	0.384***	0.388***		0.273***	0.271***	
	(0.018)	(0.027)	(0.031)		(0.023)	(0.025)	
Age at Release	-0.056***	-0.073***	-0.055***		-0.038***	-0.032***	
	(0.007)	(0.012)	(0.014)		(0.009)	(0.011)	
Drafted	0.011	0.021			-0.006		
	(0.039)	(0.060)			(0.051)		
Time since Signed	-0.030***	-0.028**	-0.049***		-0.030***	-0.037***	
	(0.008)	(0.013)	(0.015)		(0.011)	(0.012)	
Hitter	0.147***						
	(0.031)						
# of Cards	0.077***	0.086***	0.082***		0.062***	0.061***	
	(0.005)	(0.007)	(0.008)		(0.007)	(0.008)	
Experience	0.252***	0.312***	0.347***		0.189***	0.206***	
	(0.024)	(0.034)	(0.039)		(0.032)	(0.035)	
Draft Rd June Reg			0.115***			0.074***	
			(0.028)			(0.020)	
Draft Pick June Reg			-0.005***			-0.003***	
			(0.001)			(0.0008)	
Firstbaseman	0.126***	0.125***	0.157***				
	(0.037)	(0.040)	(0.046)				
Secondbaseman	-0.006	-0.010	-0.031				
	(0.036)	(0.039)	(0.045)				
Thirdbaseman	0.027	0.028	0.054				
	(0.035)	(0.039)	(0.044)				
Shortstop	-0.024	-0.036	-0.044				
	(0.036)	(0.040)	(0.047)				
Outfielder	0.102***	0.099***	0.101**				
	(0.031)	(0.035)	(0.040)				
Infielder	0.0003	-0.007	-0.011				
	(0.039)	(0.043)	(0.051)				
Designated Hitter	0.266***	0.266***	0.272***				
	(0.060)	(0.066)	(0.075)				
Constant	1.771	2.265	1.938	1.228***	1.442***	1.353***	1.046***
	(0.170)	(0.260)	(0.303)	(0.027)	(0.221)	(0.248)	(0.017)
# observations	5977	3177	2565	3177	2800	2365	2800

Table 4 – Estimation of models of final prices. Model 1 includes career statistics, Model 2 is a Tobit, and Model 3 uses only initial variables. All models include random effects by player. The letters H and P represent models for Hitters and Pitchers, respectively. Standard errors are in parentheses, and ***, **, and * represent significance at the 1%, 5%, and 10% level respectively.

	(1H)	(2H)	(3H)	(1P)	(2P)	(3P)
Black	-0.045	-0.032	-0.033	0.021	0.167	-0.089
	(0.038)	(0.074)	(0.062)	(0.053)	(0.137)	(0.080)
Hispanic	-0.058	-0.137	0.108	-0.048	-0.106	-0.028
	(0.051)	(0.101)	(0.089)	(0.050)	(0.130)	(0.079)
RC Initial	0.648***	1.551***	0.615***	0.438***	1.490***	0.388***
	(0.026)	(0.058)	(0.029)	(0.025)	(0.079)	(0.027)
XRC Initial	0.679***	1.519***	0.608***	0.413***	1.340***	0.373***
	(0.030)	(0.062)	(0.033)	(0.029)	(0.081)	(0.031)
Age at Release	-0.025**	-0.084***	-0.099***	-0.019**	-0.101***	-0.060***
	(0.010)	(0.020)	(0.017)	(0.008)	(0.021)	(0.012)
Drafted	-0.027	-0.168*	0.068	-0.019	-0.088	0.037
	(0.048)	(0.096)	(0.086)	(0.042)	(0.117)	(0.068)
Time since Signed	-0.033***	-0.115***	-0.036**	-0.019**	-0.091***	-0.012
	(0.010)	(0.021)	(0.018)	(0.009)	(0.025)	(0.014)
Experience	-0.263***	-0.245***	-0.157***	-0.206***	-0.295***	-0.127***
	(0.037)	(0.062)	(0.043)	(0.036)	(0.078)	(0.043)
# of Cards	0.001	0.013	0.099***	-0.006	0.013	0.047***
	(0.006)	(0.011)	(0.010)	(0.006)	(0.015)	(0.009)
Hits	0.002***	0.001*				
	(0.0003)	(0.0006)				
Homeruns	0.004***	0.003***				
	(0.0003)	(0.0006)				
Stolen Bases	0.001*	0.00009				
	(0.0003)	(0.0006)				
At Bats	-0.0004***	0.00001				
	(0.00007)	(0.0001)				
Postseason Games	0.009***	0.010***		0.016***	0.011*	
	(0.001)	(0.002)		(0.003)	(0.006)	
Wins				0.006***	0.001	
				(0.002)	(0.005)	
Games				-0.001***	-0.0001	
				(0.0003)	(0.0007)	
Shutouts				-0.023**	-0.048*	
				(0.011)	(0.025)	
Saves				0.003***	0.004***	
				(0.0003)	(0.0007)	
Strikeouts				0.001***	0.001***	
				(0.0001)	(0.0002)	
Constant	0.708***	1.101**	2.402***	0.541***	0.935*	1.542***
	(0.232)	(0.465)	(0.368)	(0.195)	(0.514)	(0.297)
# observations	2949	2949	3137	2604	2604	2761

Table 5 – Panel data regressions modeling change in ln(price/common) from initial pricing to 17 year later pricing. Models for hitters and pitchers are estimated separately. Standard errors are in parentheses, and ***, **, and * represent significance at the 1%, 5%, and 10% level respectively.

	Hitters	Pitchers
Black	-0.156***	-0.105
	(0.056)	(0.081)
Hispanic	-0.059	-0.031
	(0.076)	(0.076)
RC Initial	0.258***	0.079**
	(0.036)	(0.032)
XRC Initial	0.291***	0.129***
	(0.041)	(0.036)
Age at Release	0.043***	0.011
	(0.015)	(0.012)
Drafted	-0.030	-0.034
	(0.072)	(0.065)
Time since Signed	-0.012	0.007
	(0.016)	(0.014)
Experience	-0.503***	-0.312***
	(0.052)	(0.049)
# of Cards	-0.053***	-0.053***
	(0.009)	(0.009)
Hits	0.001***	
	(0.0005)	
Homeruns	0.002***	
	(0.0005)	
At Bats	-0.0004***	
	(0.0001)	
Postseason Games	0.001***	0.015***
	(0.002)	(0.004)
Wins		0.008**
		(0.003)
Shutouts		-0.030*
		(0.017)
Saves		0.002***
		(0.0005)
Strikeouts		0.0005***
		(0.0002)
Firstbaseman	-0.153***	
	(0.059)	
Secondbaseman	0.018	
	(0.058)	
Thirdbaseman	-0.091	
	(0.056)	
Shortstop	0.055	
	(0.057)	
Outfielder	-0.126**	
	(0.050)	
Infielder	0.021	
	(0.064)	
Designated Hitter	-0.310***	
	(0.099)	
Constant	-1.505***	-0.766***
	(0.345)	(0.298)
# observations	2949	2604

[1] See Zillante (2008, 2012) for more detailed information on the baseball card industry and baseball card bubble.

[2] For a small percentage of RCs, multiple players appear on the card. For instance, Jose Canseco shares a 1986 Fleer card with Eric Plunk. We follow the standard approach in the literature and omit these from our analysis.

[3] We omit 21 potential observations from 1993 Flair and 1986 and 1987 Sportflics because they do not provide player positions.

[4] In fact, because they are generally nearing retirement, prices for active players are remarkably stable at this point. To verify this choice did not affect our results, we ran regressions identical to those presented, except that a "still active" binary variable was included. The variable was positive and significant in the panel OLS regressions, but insignificant in panel tobit regressions. Results concerning race remain similar and are available upon request.

[5] Don Wakamatsu was the only player with Asian heritage. He is excluded from the analysis.

[6] The twelve omitted players are: Al Osuna, Brett Merriman, Ken Ramos, Kevin Baez, Levon Largusa, Mickey Pina, Mike Diaz, Randy Velarde, Rob Ducey, Tino Martinez, Tony Fossas, and Tyrone Hill. Other than Tino Martinez, none of these players had long careers, nor were they highly anticipated rookies. Some may wonder how Derek Jeter, a superstar who has a black father and white mother, is classified. We classify him as black, although our regression results change very little in alternative estimations where he is classified as white.

[7] We have included a table of these averages in an appendix that is available on request.

[8] The 95th-100th percentile ranked in this manner includes Hall of Fame caliber starters (Greg Maddux, Randy Johnson, Pedro Martinez, etc.) and relievers (Mariano Rivera and Trevor

Hoffman). As such, the product of innings pitched and games played does not seem to give a clear advantage to pitchers of either type.

[9] Using the natural log of price follows the majority of previous studies in this literature. In addition, it provides a better fit than using a model in which the dependent variable is the ratio of the card price to the common card price.

[10] There are nine position variables: catcher, first base, second base, third base, shortstop, outfielder, infielder, designated hitter, and pitcher. Infielder is coded one in those cases in which a card lists the player's position as "IF", rather than giving a specific assignment. Some cards list multiple positions for the player (e.g., Jim Leyritz is listed as a catcher-outfielder-third baseman). For these cases, we assign the value 1 for each of the listed positions.

[11] While all estimated models include team specific binary variables, with the Cincinnati Reds excluded as the reference group, the coefficients on these variables are excluded from the presented tables to save space. Team coefficients are available upon request.

[12] The left censoring occurs because many RCs are classified as common cards by the 17 year mark, producing a ratio of 1. Because RCs generally have above-common prices when first released, the initial price and change in price distributions do not display this censoring issue.

[13] In addition to the results presented here which have separate binary variables for black and Hispanic players, we estimated all models with a single binary variable that includes both. Our results are qualitatively similar to those for black players in the models presented in the paper.

[14] Models including only cards with RC or XRC status at the final pricing yield similar results.

[15] To separate out the effect of the initial price, we also estimated a model with the natural log of the final price as the independent variable and the natural log of the initial price as a dependent

variable. Estimated results for variables of interest remain the same as the results presented in the paper, and the coefficient estimate for the initial price variable is insignificant.