Do expected marginal revenue products for National Hockey League players equal their price in daily fantasy games?

Benny Goldman

June 19, 2014

Adviser: Professor Raymond Robertson
Macalester College

Abstract

The equality between wages and marginal revenue products is a backbone of competitive labor markets. This study will seek to test the congruity between the two in the market for players in daily fantasy hockey games. Any observed and statistically significant incongruity would lead to the conclusion that an individual can earn long run profit playing daily fantasy games. Both fixed effects and pooled regressions are employed to isolate inequalities between prices and expected marginal revenue products for players in daily fantasy hockey games. Any deviation of such could potentially be explained by utility maximizing gamblers or incomplete information. Robust results suggest that players playing at home and players playing against weak opponents relative to their own team strength are undervalued. Players who have performed above their average performance in recent games are overvalued. Although it is clear expected marginal revenue products and prices do not equate, performance is still largely random and hard to predict.

Contents

References, tables, and figures follow the conclusion of the paper.

I Introduction

Do expected marginal revenue products for National Hockey League (NHL) players equal their price in daily fantasy games? The immense popularity of professional sports in the United States has led to rapid growth in a number of secondary markets. A fairly recent development has been the emergence of fantasy sports in general, and fantasy leagues, in particular. Fantasy leagues give fans the opportunity to "draft" and trade for players, as a general manager of a sports franchise would, in order to compete against teams chosen by other fantasy owners.

Scoring in fantasy leagues is based on the performance data of real athletes in live competition. These fantasy games are traditionally played over the course of a season, beginning with a draft. The players chosen in the draft are yours to keep for the ensuing season (barring any trades or additions of players not chosen in the draft). The popularity of fantasy games in the last decade has led to a new format: the daily game.

The daily fantasy games give avid players an opportunity to assemble a new team and compete on a daily basis. The daily games have markets for players. You are given an artificial budget of $\$55,000$ and asked to select a roster of nine players, each of whom range in price from $\$3,000$ to $\$12,000$.

The market for NHL players in daily fantasy games may be partially composed of consumers seeking to maximize utility instead of points. In a competitive market with perfect information the value of all players would be equal. In other words, the cost of a fantasy point would be equal across all players. The standard deviation of the cost of a fantasy point based on player season averages is $191.20. The Calgary Flames' Mark Giordano can get you a fantasy point for $\$1,264.71$, but if you want a point out of Jeff Carter of the Los Angeles Kings that is going to run you $\$3,850$. If you believe that season averages are a decent predictor of daily performance (this hypothesis will be tested in this paper), then you believe there is money to be made by playing daily hockey games.

The goal of this paper will be to analyze the relative efficiency of the player markets and determine whether there is potential to make long run profit as a team owner in a daily fantasy game. This analysis will be done using the point structure and player prices from FanDuel, a leader in the daily fantasy game industry. According to FanDuel's "about" page, it pays out

\$6,000,000 in prizes every week. Some players, it is claimed, gross \$5,000 a day in winnings (https://www.FanDuel.com/about-us).

In section 2 a review of the literature on hockey performance and marginal revenue products is presented. Section 3 will build a theoretical framework for analysis. Sections 4, 5, and 6 will include summary statistics, analytics, and robustness checks. In section 7 conclusions are drawn and directions for future research are discussed.

II Literature Review

Any deviation in prices from expected marginal revenue products could be due to utility maximizing behavior (instead of fantasy point or profit maximizing behavior) on the behalf of consumers or incomplete information. Paul and Weinbach (2010) looked closely at the forces that drive consumer behavior with respect to sports gambling. They found that investment-based gamblers are in the minority and that consumers' betting decisions are often determined by which team is being broadcast on television or which team is believed to be "better." Both team quality and television availability had positive and significant effects on betting volume. These types of biases among consumers can lead to inefficiencies in betting lines, and in the case of daily fantasy games, player prices. If players who are going to be playing in a nationally broadcast game see an increase in demand, they become overvalued and priced above their anticipated productivity levels.

Inefficient pricing, defined as deviations from the price called for by competitive markets with perfect information, are certainly not unique to daily fantasy games. After all, it was mispricing that launched the Moneyball revolution in Major League Baseball. Oakland Athletic's general manager Billy Beane used statistics like slugging percentage and on base percentage to value prospects (as opposed to accepted measures like batting average and runs). Hakes and Sauer (2006) confirmed that on base percentage was indeed undervalued in the market for baseball players. As a result, the Oakland Athletics were able to thrive in Major League Baseball's American League despite a consistently small pay roll. This performance boost lasted only a short while. Hakes and Sauer (2007) determined that the benefits of the strategy are largely dependent on the number of imitators. Once teams processed the value of on base percentage, prices adjusted, and the advantage was all but gone.

Billy Beane's success in Oakland was well documented. Author Michael Lewis wrote a book titled "Moneyball" and soon after teams in all four major sports began to employ advanced statistical analytics to improve in game strategy and prospect valuations. Mason and Foster (2007) concluded that the implications of Moneyball to hockey might be limited. Baseball play is isolated; the batter and pitcher exist in a near vacuum. Much of a player's hockey performance, on the other hand, is a function of the talent around him. Predicting and evaluating hockey performance is a challenge because there are limited statistics available. That being said, after the implementation of the salary cap in the NHL, the valuation of prospects became all the more important and teams started to open up to the idea of using statistical techniques to value players (Mason et al., 2007).

The first step required to understand the value of an NHL player is determining the variables that contribute to performance. Gramacy et al. (2012) argues that the plus-minus statistic is flawed. The plus-minus is a common statistic used to measure hockey performance. You earn a point for being on the ice when your team scores an even strength goal and you lose a point for being on the ice when the opposing team scores an even strength goal. Gramacy et al. (2012) notes that this only measures marginal effects of players and becomes a rather inaccurate predictor of performance because it does not control for skill of your teammates or your opposition. Gramacy et al. (2012) use logistic regressions and conclude that most players do not have a "measurably strong" effect on team performance. This supports the belief that the NHL is a star-driven league. The small variability in player effect on team performance allows the stars to look especially great and leads to the existence of undervalued prospects (Gramacy et al., 2012). Not only do undervalued prospects exist, but "some of the higher paid players in the league are not making contributions worth their expense," (Gramacy et al., 2012).

Beyond the plus-minus, teams have traditionally used other statistics in an attempt to value performance. Kahane (2001) concludes that much of the variability in NHL player pay can be attributed not only to differences across players, but differences across teams. Kahane (2001) uses a maximum likelihood estimator to demonstrate the difference across teams. Kahane (2001) estimates that 2.2% of variability in player salaries can be attributed to the fact that different teams have different willingnesses to pay, mostly due to varying levels of revenue. Teams also differ in how they reward changes to performance. Certain teams reward or punish deviations from expected production more so than others (Kahane, 2001).

Much of the variability in pay among players can be explained with a few basic hockey statistics. Points per game, all star game appearances, penalty minutes per game, and being picked in the first or second round of the draft all had positive and statistically significant effects on pay (Eastman et al., 2009). Eastman et al. (2009) also found that the plus-minus statistic was a strong determinant of earnings, specifically for defenseman. Eastman et al. (2009) reported heteroskedasticity as an estimation issue. It turns out that career statistics are far more predictive of pay for high paid stars than for low paid stars.

The literature on hockey performance and its relation to player value is very focused on determinants of career performance and how those affect teams' willingnesss to pay a player. Instead of focusing on teams and seasons, this paper will seek to fill a niche in the literature by determining the variables that affect daily performance of NHL players and then testing the efficiency of daily player prices.

III Theory

In order to build a theoretical framework to understand pricing deviations from marginal revenue products, the factors that drive expected marginal products must be established, and then it must be determined if those same variables affect prices. It will be assumed that players of daily fantasy games are attempting to maximize points subject to choosing nine different players and constrained by the $55,000 (of FanDuel money) salary limit. This assumption will be relaxed later.

Let X_1 indicate player 1, X_2 indicate player 2...all the way to player X_n where n is the number of possible players. These are dummy variables that take the form of 1 (active) or 0 (not active) based on whether the player is chosen by the daily fantasy owner. Let S equal the budget of each daily team owner (the $55,000 of "FanDuel" money given to be spent on players). M_1 is the expected marginal product of labor for player X_1, M_2 is the expected marginal product of labor for player X_2...M_n is the expected marginal product of labor for player X_n. Therefore the points a daily fantasy game owner can expect to score are given by:

$$ExpectedPoints = M_1X_1 + M_2X_2 + M_3X_3... + M_nX_n$$

To maximize points, the fantasy owner will need to use the full budget.

$$S = P_{X_1}X_1 + P_{X_2}X_2 + P_{X_3}X_3... + P_{X_n}X_n$$

The maximizing decision can be expressed as:

$$\underset{X_{1-n},\lambda}{max} \quad \Phi = (M_1X_1 + M_2X_2 + M_3X_3... + M_nX_n) - \lambda(S - P_{X_1}X_1 - P_{X_2}X_2 - P_{X_3}X_3... - P_{X_n}X_n)$$

First order conditions 1, 2, and $n+1$:

$$\frac{\partial\Phi}{\partial X_1} = M_1 + \lambda P_{X_1} = 0 \quad (1)$$

$$\frac{\partial\Phi}{\partial X_2} = M_2 + \lambda P_{X_2} = 0 \quad (2)$$

$$\frac{\partial\Phi}{\partial\lambda} = -S + P_{X_1}X_1 + P_{X_2}X_2 + P_{X_3}X_3... + P_{X_n}X_n = 0 \quad (3)$$

First order conditions one and two can be rearranged such that:

$$M_1 = -\lambda P_{X_1} \quad (4)$$

$$M_2 = -\lambda P_{X_2} \quad (5)$$

Equations 4 and 5 can be combined to show the optimization decision of the daily fantasy player.

$$\frac{-\lambda P_{X_1}}{-\lambda P_{X_2}} = \frac{M_1}{M_2}$$

The above equation can be simplified and rearranged.

$$\frac{P_{X_1}}{M_1} = \frac{P_{X_2}}{M_2}$$

A profit-maximizing individual will maximize points by setting the ratio of the price of the player to his expected productivity equal across all players. Because individuals face the same prices and expected points, these ratios would equate across the market. This would mean that all individuals are indifferent to the players they choose so long as the whole $\$55,000$ budget is used. In such a scenario, only risk loving or risk neutral individuals would choose to participate in daily fantasy games.

To demonstrate this concept, imagine that player 1 and player 2 have prices and expected points such that $\frac{P_{X_1}}{M_1} > \frac{P_{X_2}}{M_2}$. This would mean that a point from player 1 is more expensive than a point from player 2. An individual who is attempting to maximize points on their fantasy team would never choose player 1 over player 2.

The demand for player 1, M_1, is perfectly elastic. Any price above P_1 would exceed the expected marginal product. No profit maximizing fantasy player would choose player 1 if that were the case. Any price below P_1 would make player 1 undervalued. His expected marginal product exceeds his cost. In such a scenario, all profit maximizing individuals would select player 1. This would put upward pressure on price which would continue to rise until it returns to P_1.

As Paul and Weinbach (2010) suggest, individuals may not be making optimal decisions with respect to sports gambling. Many are influenced by exterior variables such as television availability, or individuals do not have complete information. This would suggest a departure from perfect competition as some individuals would be willing to pay above a player's marginal product if that player's game is on television.

For example, assume player 1 is playing in a nationally broadcast game. Individuals who receive utility from selecting player 1 and then watching him on television would have a willingness to pay for player 1 above P_1. This would put upward pressure on price, driving P_1 above M_1. These utility-maximizing individuals will now be selecting an overvalued player. As a result of the increase in demand for player 1, there is less demand for the remaining space of players (X_{2-n}). This would put downward pressure on price and result in markets where the marginal product exceeds the price. The increased demand for player 1 causes the rest of the players to be undervalued. Individuals who are profit maximizing and do not choose player 1 will score more points than those who do choose player 1. The presence of utility maximizing behavior creates potential long run profit for daily game players.

Uncertainty may also result in a departure from perfect competition. Because prices are pre-determined and displayed for all individuals, there is no uncertainty with respect to prices. That being said, the values of M_{1-n} are not known. In order to maximize points, individuals must first make a prediction or a probabilistic distribution of the values of M_{1-n}. Those who are able to make the most accurate predictions of M_{1-n} will have the highest probability of maximizing points and an opportunity to earn long run profit.

The remainder of this paper will analyze whether or not expected marginal revenue products do in fact equal prices. Any deviation between the two leaves an opportunity for individuals to make a long run profit by selecting undervalued players. To do this, two equations must be estimated.

First, it must be determined what variables affect player performance on a night-to-night basis. The quality of a player, his recent performance, his health, his opponent's fatigue, whether or not the player is playing at home, and the relative strength of his team compared to his opponent would all be expected to have a positive correlation with fantasy points. The fatigue of a player is hypothesized to have a negative correlation with fantasy points.

The following will be the guiding equation.

$$\begin{aligned} FantasyPoints_{it} =& \alpha + \beta_1 PlayerQuality_{it} + \beta_2 RecentPerformance_{it} + \beta_3 Fatigue_{it} + \\ & \beta_4 StrengthDifferential_{it} + \beta_5 OpponentFatigue_{it} + \beta_6 HomeIce_{it} + \beta_7 Health_{it} + \varepsilon_{it} \end{aligned}$$

The next step will be to determine if the variables that affect performance are the same variables that affect price.

$$\begin{aligned} Price_{it} =& \alpha + \beta_1 PlayerQuality_{it} + \beta_2 RecentPerformance_{it} + \beta_3 Fatigue_{it} + \\ & \beta_4 StrengthDifferential_{it} + \beta_5 OpponentFatigue_{it} + \beta_6 HomeIce_{it} + \beta_7 Health_{it} + \varepsilon_{it} \end{aligned}$$

Although it may seem obvious that the above variables are determinants of performance, correctly estimating the magnitude of each effect as well as assembling the optimal team requires fantasy owners incur a large time cost. This time cost may help explain any potential incongruity between expected marginal products and prices.

If it becomes clear that there are variables that affect performance, but not price, then prices and expected marginal revenue products are not equal. The presence of variables that affect price, but not performance, would also suggest an inequality between prices and marginal revenue products.

If price and performance are determined by the same variables and magnitudes, then marginal revenue products and prices are equal.

IV Summary Statistics

Theory states that all variables that may affect a hockey player's performance for a given game must be present in both the regression on price and the regression on fantasy points. These are measures of player quality, recent performance, fatigue, team strength, opponent strength, opponent fatigue, and whether or not the player is playing at home.

Panel data were collected from the hockey statistics website Hockey Reference (http://www.hockey-reference.com). Hockey Reference uses the official statistics provided by the NHL and has game by game data dating as far back as 1917 (http://www.hockey-reference.com). For sake of consistency, performance statistics were only collected on players for which there were also data available on their FanDuel market prices. There is no archive of FanDuel market prices, and the data therefore had to be collected daily. The collection began on January 9, 2014, and continued for approximately a month until the NHL Olympic break on February 8, 2014. The resulting dataset contained 2240 observations of FanDuel market prices on 201 different players.

Data were collected for every game in the 2013-2014 NHL season for the 201 players. A summary of the data collected appears in Table 1. Only games where there are observations of the player's price are used in analysis. The collection on prices began on January 9, 2014. This is about halfway through the NHL season, and the data have a few helpful qualities as a result. As seen in Figure 1, the winning percentages of the NHL teams are very stable by the beginning of January. This means that measures such as team strength and opponent strength are more accurate than they would be earlier in the season. The same goes for the player's season averages. The season averages are more representative of the player's ability as the season goes on.

The first dependent variable, fantasy points, are defined by FanDuel to be:

$$
\begin{aligned}
FantasyPoints = & 3Goals + 2Assists + PlusMinusRating + .25PenaltyMinutes + \\
& .5PowerPlayGoals + .5PowerPlayAssists + .4Shots
\end{aligned}
$$

Observations of fantasy points were between -5 and 14.60. Fantasy points had a mean of 2.08 and a standard deviation of 2.58 points. Clearly, player performance is very variable with a standard deviation that is larger in magnitude than the mean. There were 2240 observations of fantasy points, an average of 11.2 observations per player (meaning there was about 11 games of data for each of the 201 players). The calculation of fantasy points serves as an all encompassing performance variable. It includes all of the relevant statistical measures of hockey performance, and weights them according to their importance.

The second dependent variable, FanDuel market prices, are posted daily on FanDuel. The 2240 observations ranged from 3000 to 10500 with a mean of 4747.77 and a standard deviation of 1395.94. The large variability in prices complements the large variation in performance.

The player quality variable can be treated in one of two ways. In the main results section of the paper (Table 2) a fixed effects regression is used to control for time invariant differences across players. This is the most accurate way to control for player quality because it is effectively generating a dummy variable for each player. A second method (which will also be tested) is to use the player's season average fantasy points as a proxy variable for player quality. The reason this method is inferior is that it becomes more accurate and less variable as the player has more observations. The season averages have a mean of 2.35 with a standard deviation of 0.65, a minimum of 1.20, and a maximum of 4.88. These can be compared to the daily measurements of fantasy points, which are far more variable. A player's performance in a given game may have a large random component, but better players have higher averages when large samples of games are pooled. Figure 2 shows the kernel density plots of season average fantasy points and daily fantasy points.

Recent performance is measured in a few different ways. The first is simply an average of the player's fantasy points in his last 5 games. This measure ranged from -0.74 to 8.20 with a mean of 2.23 and a standard deviation of 1.37. Using the same technique but including only the last 3 games yields a mean of 2.22, a standard deviation of 1.69, and a range from -1.50 to 9.57. This continues to support the notion that performance is more volatile over short samples of games. The 5 games metric has a lower standard deviation. A second set of recent performance variables for which the player's season average is subtracted from their performance in the last 3 and 5 games was also created. The moments of these transformed variables are given in Table 1.

The team strength differential variable measures how strong a player's team is relative to his opponent. The variable was calculated by simply subtracting opponent winning percentage from team winning percentage (both of these variables are also shown in Table 1). The strength differential variable ranges from -.44 to .44 with a mean of 0.03 and a standard deviation of 0.16. Negative values suggest a team is playing a stronger opponent, and positive values are given for teams playing a weak opponent relative to their own team.

Home ice, fatigue, and opponent fatigue are all dummy variables. Whether or not a player or opponent played in a game yesterday are used as proxies for fatigue and opponent fatigue. The home ice variable takes on a value of 1 if the player is playing home and a value of 0 if the player is on the road. The mean is .51 and the standard deviation is .50. The played yesterday variable is 1 if the player had a game yesterday and 0 if the player did not. It has a mean of .17 and a standard deviation of .37. Similarly, the opponent played yesterday variable is 1 if the opposing team had a game the previous day and 0 if they did not. The mean of the opponent played yesterday variable is .16 with a standard deviation of .37. Hockey games are usually not scheduled on back-to-back days, hence the low means of the played yesterday and opponent played yesterday variables.

V Analysis

V.1 Estimation Issues

The first issue is multicollinearity between the regressors. This is only an issue between the variable used to control for player quality and the measures of recent performance. A player who has a high season average is likely to have a similarly high average for performance in their last 5 games. This is because the player's performance in the last 5 games is included in the season average. Season average and performance in the last 5 games have a correlation coefficient of .53. Season average and performance in the last 3 games have a correlation coefficient of .42. The variance inflation factors on season average and the average of performance in the last 3 games are 1.28 and 1.27, respectively. Although the variance inflation factors are low, the large correlation coefficients are still cause for concern.

By simply differencing the player's season average from his recent performance, the variance inflation factors on both variables fall to 1.01. Not only do the variance inflation factors fall, but also the interpretation of the coefficients is now straightforward. The coefficient can be interpreted as the effect on fantasy points of a one point increase in the difference between recent performance and season average. The demeaning of the season averages allows for easy comparison across players of different quality.

Heteroskedasticity is present in all tested specifications. Figure 3, a plot of the combined residuals (fixed component and the overall component) against the fitted values for the main estimation equation (Table 2 column 3), has the appearance of homoskedasticity. A modified Wald test (which has a null hypothesis of homoskedastic data) on the fixed effects regression yields a P value of 0.0000. This would lead to a rejection of the hypothesis that the data is homoskedastic.

Due to the unavailability of the proportionality factor, a weighted least squares method to remedy the heteroskedasticity is not feasible. Not only is the proportionality factor unobserved, but weighted least squares also complicates the coefficient interpretations. Instead, I report robust standard errors that account for the probable downward bias on the standard errors caused by heteroskedasticity.

A third potential issue is the relative accuracy of the proxy variables used to capture fatigue. Because team practice and travel schedules are not made public, it is difficult to estimate the fatigue of one player compared to another. While playing yesterday certainly would induce fatigue, not playing yesterday does not mean a player is fresh. Often teams have practices or spend their off days traveling, both of which can be close to as draining as playing a game.

Similar to fatigue, injuries are also likely to affect performance, but information on injuries is not made available to the public unless the injury is extreme. Some teams go the full season without ever listing a player on the injury report, and nearly all injury reports are for players that have been placed on injured reserve and will be forced to sit out. Unlike the National Football League, the National Hockey League does not require teams to submit a full list of their injured players and the severity of the player's injuries before each game. The lack of public injury reports is a potential estimation problem. Hockey is an extremely violent sport, and it is not realistic to assume that the only players bothered by injury are the very few (if any at all) who the teams list on their injury reports.

The danger of an inaccurate proxy for fatigue and the lack of an injury variable is endogeneity.

If injuries affect fantasy points on a given night, the error term would now include the effect of the injuries. Endogeneity would then be present if injuries are correlated with any of the regressors. This does not appear to be the case. A regression of the residuals on the independent variables results in coefficients that approach 0 and t statistics of 0.00 (and P values of 1.000). The lack of an injury variable and a potentially poor proxy for fatigue do not appear to cause endogeneity.

V.2 Main Results

The main results are presented in Table 2. The first step in examining the relative efficiency in the market for players is pinpointing the determinants of the marginal products, and then the determinants of the prices. If these two are determined by different variables, then undervalued and overvalued players exist in the market. The results presented show that the determinants of the marginal products and prices are in fact different.

Given the name of a player, it would be very tough to predict that player's performance on a given night. The opposite is true for prices. A simple regression of price on a dummy variable for each of the 201 players yields an R-squared of .96 and an F test with a P statistic of 0.000. This means that 96% of the variation in prices can simply be explained by the players, and prices are statistically significantly different across players. A similar phenomenon does not exist for the marginal products. A regression of fantasy points on a dummy variable for each player has an R-squared of 0.14 and an F test with a P statistic of 0.000. Although marginal products are statistically significantly different across players, differences across players only account for 14% of the variation in fantasy points.

The next step is to find the determinants of fantasy points and also test if players are truly the only variable that prices are dependent upon. As determined by theory, a control for player quality is relevant in estimating fantasy points. There are two methods of controlling for player quality. The first involves using a proxy variable, season averages, that would control for differences in player skills by differentiating players via their average production. The potential issue with season averages is that they are more variable in the beginning of the season when the player has played fewer games, and they become a more accurate measure of quality as the season goes on.

A second method of controlling for player quality is simply estimating a fixed or random effects regression. The random effects regression would be appropriate if the unobserved differences in player quality are uncorrelated with the regressors, whereas the fixed effects regression would be appropriate if correlation between the time invariant differences between players and the regressors existed. A Hausman test with a null hypothesis of no systemic differences in coefficients was employed to determine which regression was appropriate. The test returned a P value of 0.0000. This would lead to a rejection of the null hypothesis that a random effects regression is appropriate. The fixed effect regressions were estimated on both fantasy points (column 3) and price (column 4). The regressions were estimated with the same independent variables, but the results were quite different. Robust standard errors are reported in order to correct for downward bias of the standard errors caused by heteroskedasticity.

Theory predicted that a player who is performing in his home arena would have a higher marginal product than a comparable player who is playing an away game. The regression in column 3 confirms this result. Holding team strength, opponent strength, recent performance, fatigue, and opponent fatigue constant, being at home increases a player's production by 0.39 fantasy points. The result is statistically significant at the 0.01 level. The same result, however, is not present in the regression on price. Playing at home has an extremely small and non-statistically significant effect on price. Home players are undervalued in the market. Given a choice between two players who only differ in game location, the daily fantasy owner can increase their expected points by selecting the player who plays at home (it does not, however, appear that all fantasy owners are doing that).

It was also projected that players would score fewer fantasy points against a strong opponent than they would against a weak opponent. The regression supported that hypothesis. For a one unit increase in the difference between a player's team strength and his opponent's strength, that player can be expected to score .98 more points. A one unit increase is not realistic, however, because the winning percentages range from 0 to 1. An increase by 0.16 units (the standard deviation) would be expected to increase a player's output by .157. The result was statistically significant at the 0.05 level. Opponent strength relative to team strength was not a statistically significant determinant of price. The standard error (50.40) was nearly five times as large as the estimated coefficient (12.64). The coefficient, 12.64, was also very small when taking into account the average player price is 4747.77. This means that players who are matching up against strong teams relative to the strength of their own team are overvalued on the market. Players matching up against weak teams relative to their own team are undervalued on the market, and avoiding the selection of these players can lead to long run profit.

Contrary to what was predicted by theory, the recent performance of a player had a negative and statistically significant effect on fantasy points. The effect, however, was not economically significant. The regression predicted that holding all else constant, a player who increases his recent performance in the last 3 games over his season average by one point would experience a decline in fantasy points of .17. The result was significant at the 0.01 level. The effect of recent performance on price actually had the opposite effect. "Hot" players seem to become more popular and apply upward pressure on price. A one unit increase in the average performance of a player in his last 3 games over his season average yields an increase in price of 85.44. The estimated parameter was statistically significant at the 0.01 level. Players who have performed above their season averages in the last 3 games are overvalued. There is a close to 0 change in expected fantasy points, but a sizeable increase in price for a player that has been "hot."

Neither playing yesterday nor facing an opponent who played yesterday had a statistically significant effect on fantasy points. Both estimated coefficients were negative and small. Theory predicted that if playing yesterday is an accurate proxy of fatigue, players who played yesterday would score fewer points and players who are facing an opponent who played yesterday would score more points. Although the effect was not statistically significant from 0, an opponent playing yesterday caused a statistically significant (at the 0.01 level) drop in price of 40.62. A player who played yesterday did not, however, have a statistically significant effect on price. This suggests that players who are facing an opponent who played yesterday are improperly valued, and the market believes players will score less points if their opponent had a game yesterday.

The results emphatically support the notion that marginal revenue products do not equate to prices in the market for players. Players who are playing at home and playing against a team that is weak relative to their own are undervalued. Players that have performed above their season average in their last 3 games are overvalued. Lastly, players facing an opponent who played yesterday are undervalued due to a drop in price with no observable effect on their performance. Paul and Weinbach (2010) found this same effect, and determined that people sometimes make betting decisions based on factors that do not affect outcomes.

Expected performance is still relatively unpredictable. Much of the variation in prices can be explained by the player, but only 16% of the variation in fantasy points can be explained by the variables presented. This characteristic will be further explored in the section to follow.

V.3 Robustness

This section will offer alternative approaches to the estimation methods presented in the main results table in order to test the sensitivity of the parameters to various specifications and regressions. Table 3 uses a pooled regression approach, as opposed to a fixed effects regression, in order to control for differences in player quality. Also in Table 3 (columns 3 and 4), a near identical regression to those presented in the main results table is used, but a five game standard to measure recent performance is implemented (instead of the 3 games standard used in Table 2). Table 4 uses fixed regressions with controls for first order autoregressive disturbances. This method, originally presented by Baltagi and Wu (1999), is useful to estimate parameters of unbalanced panel data where observations are not equally spaced over time or equally frequent across actors. The method is also helpful in controlling for issues caused by heteroskedasticity (which is present). Table 5 estimates a random effects regression. Table 6 returns to the original fixed effects estimation method, but makes substitutions for the team strength differential and recent performance variables. Lastly, Table 7 uses pooled regressions with season averages to assess the sensitivity of the variable substitutions made in Table 6.

The results presented in Table 2 all use a fixed effects regression to control for player quality, or time invariant differences across players. Table 3, columns 1 and 2, present similar regressions to those presented in Table 2, but they instead use season averages as a proxy for player quality instead of a fixed effects regression. The results only slightly differ from those found in Table 2. Robust standard errors are reported in Table 3 as well do to the persistence of heteroskedasticity in the pooled regression. The Breusch-Pagan/Cook-Weisberg test for heteroskedasticity (null hypothesis of homoskedastic data) returns a P value of 0.000 for all four regressions in Table 3.

Both the fixed effects regression and the pooled regression with season average controls (Table 3, column 1) yield positive and statistically significant effects of playing at home. The size of the effect is not statistically different between the two regressions. Similarly, in both regressions the effect of playing home on price is not statistically significantly different from 0. Players who are playing at their home arena are still under valued when season averages are used as a proxy for player quality.

The parameter estimate on season averages are positive and statistically significant on both fantasy points and price. A one unit increase in a player's season average would mean the player is predicted to score .76 more points in a given game and cost 2005.62 more (both are significant at the 0.01 level).

The estimate on the effect of opponent strength relative to team strength becomes statistically insignificant on performance, and remains insignificant on price. Using season averages to control for player quality diminishes the returns to playing a poor opponent relative to your own team. The effect falls from 2.12 to -.25 (and loses its significance). The effect on price is still small and insignificant with a standard error roughly 3.5 times the size of the coefficient estimate. A pooled specification would suggest that fantasy owners are behaving optimally with respect to their treatment of team strength differentials.

The previous economic insignificant effect of recent performance is now no longer statistically different from 0. The effect remains positive on price, but it is not quite as large. It dropped from 85.44 to a 26.51 increase in price for every one unit increase in the recent performance variable. Players who have performed above their season average in the previous 3 games remain overvalued.

Lastly, coefficients on playing yesterday and the opponent playing yesterday were both economically and statistically insignificant on fantasy points and price. This is a departure from the main results where an opponent playing yesterday put downward pressure on a player's price. Under a pooled specification, the market appears to correctly estimate the effect of playing yesterday or facing an opponent who played yesterday on performance.

Using season averages as a proxy control for player quality yields approximately the same results as the fixed effects regression in Table 2. The largest difference is the diminished effect of playing a weak opponent relative to a player's own team, and a lack of price adjustment for opponents who played the day before. The explanatory power of both models, fantasy points and prices, fell when the pooled regression was employed.

Similar to the fixed effects regression, a pooled regression with dummy variables for each player was implemented in columns 3 and 4 of Table 3. The big difference between the two methods is the use of a 5 game lag on recent performance as opposed 3 games. This style of regression actually provides an equally good fit for the data as the main results regression. The R-squared (and adjusted R-squared) on both the fantasy points regression and the price regression improve over those in Table 2.

Although the overall fit of the data improves, there are few differences in the parameter estimates between the two methods. Home ice remains positive, economically significant, and statistically significant on fantasy points, but has no significant effect on price. The team strength differential is estimated to be roughly the same size and significance as the fixed effect regression while still having an insignificant and statistically indifferent from 0 effect on price.

Including a recent performance metric based on the previous 5 games has a dramatic effect. The variable suggests that recent performance and current performance are negatively related. A one unit increase in the player's season average subtracted from the average of the player's last five games' performance leads to a drop in fantasy points of 0.28 (significant at the 0.01 level). It has an even larger effect on price than the effect estimated in Table 2. The parameter nearly doubles in size from 85.44 (from the main results) to 151.47, both of which are statistically significant at the 0.01 level. The effect of playing yesterday still remains economically and statistically insignificant on both price and fantasy points. The parameters appear to be insensitive to the removal of the opponent played yesterday variable. The exclusion of which failed to cause any disturbance.

Employing a pooled regression with a recent performance variable dating back to 5 games and dummy variables for all players results in the same conclusions as presented by a fixed effects regression. The market continues to undervalue or overvalue the same attributes, but lagging the recent performance variable back to 5 games appears to double the effect that performing above season average has on price.

Returning to the fixed effects specification used in Table 2, but employing a control for an auto regressive disturbance of the first order, yields large differences in the parameters. Although serial correlation is not present, the regression control also assists with heteroskedasticity (which is present) and is appropriate for unbalanced panel data. This estimation method yields improved explanatory power for variation within each player, but has close to 0 explanatory power for variation between players (R-squared values lower than .001).

The effect of home ice and strength of opponent relative to a player's own team become statistically insignificant with respect to both price and fantasy points. Player's who have performed above their season average in the past three games appear to be even further over valued. Recent performance has a negative effect on fantasy points but a positive effect on price (both statistically significant). Playing yesterday now has a negative and statistically significant effect on performance, and no effect on price. Facing an opponent whop played yesterday is insignificant on both.

The results are only slightly sensitive to changes in specification. Using the same auto regressive correction as just described, but dropping the played yesterday variable, dropping the opponent played yesterday variable, and replacing the 3 game lag of recent performance with a 5 game lag, leads to slightly different conclusions. Also, the effect of playing at home and playing a bad team become positive and statistically significant, as they were in the main results, but continue to have a small and statistically insignificant effect on price. The 5 game lag appears to again have a greater pull on price and performance than the 3 game lag. In Table 4 columns 3 and 4, a one unit increase in average performance in the last 5 games minus a players season average would lead to a drop in production of 1.88 points, but a rise in price of 93.82 (both statistically significant at the 0.01 level). Again, the market appears to be severely overvaluing players who have performed above their standard levels in previous games.

Table 5 uses a random effects estimator, as opposed to a fix effects estimator, with the same independent variables as used in the main results section. The results are nearly identical, except for the estimate of opponent strength relative to a player's team strength. This result becomes statistically and economically insignificant. The explanatory power of both models fall. The random effects regression can only explain .1% of the variation in fantasy points and 19% of the variation in prices.

In Table 6, a fixed effects equation was used. The regressions differ from the main results because the opponent's winning percentage is used instead of the strength differential. This removes the effect of a player's own team and isolates the opponent. Instead of using a demeaned performance in the last 3 or 5 games, the recent performance variable simply evaluates the player's average fantasy points in the last 3 or 5 games. The results stick closely to those of the main results. The only significant change is the statistical insignificance of the opponent strength variable. Adjusting the specification to remove a player's own team's quality reveals a relationship of the same sign as the main results, but with no statistical significance.

Table 7 repeats the regressions in Table 6, but uses season averages as a proxy for player quality instead of estimating a fixed effects regression. The conclusions drawn from the results are the same, except for the variables on recent performance. Using averages of the last 3 games and the last 5 games becomes insignificant when a pooled regression is used instead of a within effects estimator. The positive and statistically significant effect on price remains.

VI Conclusion

It is clear that markets for players in daily fantasy games do not equate prices to expected marginal revenue products. The mistakes in estimations of the expected marginal products are relatively consistent over various specifications and types of estimation methods. Different specifications slightly adjusted market adjustment to facing a fatigued opponent, and the performance returns to performing well in recent games and playing a weak opponent (relative to a player's own team).

There is a set of patterns that persisted through all estimation methods. Players who are playing at home score more fantasy points, but do not cost more. The market is undervaluing players who are playing in their home arena and overvaluing visiting players. The strength of a player's opponent relative to their own team has a large effect on performance, but no observable effect on price. Playing a poorer team leads to economically significant increases in performance and undervaluation in the market. Lastly, above season average performance in recent games has an effect on fantasy points that is not statistically differentiable from zero, but a large, positive, and significant effect on price. Players who have performed very well in the past few games relative to their usual performance are highly overvalued. In fact, in some specifications this actually had a negative effect on performance and a positive effect on price, increasing the market error.

While the results show that the independent variables have statistically significant effects on performance, the explanatory power of the models is limited. The highest observed R-squared was below .25, and many models could not explain 10% of the variation in performance. Figure 2, which compares the distributions of season average performance to daily performance, shows the variability in daily performance. Prices are relatively stable for each player, and there are opportunities to capitalize on variables that have significant effects on performance but not price. Due to the randomness of daily performance, however, profit is far more feasible in the long run.

The main concern with the results presented is the persistence of heteroskedasticity across all specifications and estimation methods. Robust standard errors were reported in an attempt to avoid committing a type I error. Unavailable data on player health leaves open the possibility of an endogenous error term, but all tests point endogeneity not being an issue. Also, the use of whether a player or team played yesterday as a proxy variable for fatigue is potentially improper. Travel schedules and practices can make off days equally as tiring as playing a game.

Studying the relative efficiency of season long (as opposed to daily) fantasy markets would be an appropriate follow up to this study. As seen in Figure 2, season performance is far less variable then daily performance, and if markets fail to process that, it is possible there is an even larger opportunity for profit playing season long fantasy games. A different set of independent variables, including information on age and position, may be relevant in such a study.

References

Baltagi, B. H., and P. X. Wu. (1999). Unequally spaced panel data regressions with AR(1) disturbances. Econometric Theory 15: 814823.

Gramacy, R. B., Jensen, S. T., & Taddy, M. (2013). Estimating player contribution in hockey with regularized logistic regression. Journal of Quantitative Analysis in Sports, 9(1), 97-111.

Hakes, J. K., & Sauer, R. D. (2006). An economic evaluation of the moneyball hypothesis. Journal of Economic Perspectives, 20(3), 173-185.

Hakes, J. K., & Sauer, R. D. (2007). The Moneyball Anomaly and Payroll Efficiency: A Further Investigation. International Journal of Sport Finance, 2(4).

Kahane, L. H. (2001). Team and player effects on NHL player salaries: A hierarchical linear model approach. Applied Economics Letters, 8(9), 629-632.

Mason, D. S., & Foster, W. M. (2007). Putting moneyball on ice? International Journal of Sport Finance, 2(4), 206-213.

Paul, R. J., & Weinbach, A. P. (2010). The determinants of betting volume for sports in North America: Evidence of sports betting as consumption in the NBA and NHL. International Journal of Sport Finance, 5(2), 128-140.

Vincent, C., & Eastman, B. (2009). Determinants of pay in the NHL. A quantile regression approach. Journal of Sports Economics, 10(3), 256-277.

Table 1: Summary Statistics

Variables	Observations	Mean	Standard Deviation	Minimum	Maximum
Team Winning Percentage	2,240	0.52	0.11	0.26	0.73
Opponent Winning Percentage	2,240	0.49	0.11	0.26	0.73
Team Strength Difference	2,240	0.03	0.16	-0.44	0.44
FanDuel Price	2,240	4,747.77	1,395.94	3,000	10,500
Moving Average of Fantasy Points	2,240	2.35	0.65	1.20	4.88
Fantasy Points	2,240	2.08	2.58	-5.00	14.60
Fantasy Points in Last 5 Games	2,240	2.23	1.37	-0.74	8.20
Fantasy Points in Last 3 Games	2,240	2.22	1.69	-1.50	9.57
Performance Deviation (Last 5 Games)	2,240	-0.13	1.16	-3.16	4.23
Performance Deviation (Last 3 Games)	2,240	-0.13	1.53	-4.22	7.10
Played Yesterday	2,240	0.17	0.37	0.00	1.00
Opponent Played Yesterday	2,240	0.16	0.37	0.00	1.00
Home Ice	2,240	0.51	0.50	0.00	1.00
Won Last Game	2,240	0.50	0.50	0.00	1.00
Home Ice Last Game	2,240	0.51	0.50	0.00	1.00

Table 2: Main Results

Variables	(1) Fantasy Points	(2) Price	(3) Fantasy Points	(4) Price
HomeIce			0.39***	-2.22
			(0.14)	(12.62)
TmStrengthDif			.98**	12.64
			(0.49)	(50.40)
PerformanceDev3			-0.17***	85.44***
			(0.04)	(6.34)
PlayedYesterday			-0.17	11.97
			(0.15)	(12.69)
OppPlayedYesterday			-0.03	-40.62***
			(0.18)	(15.39)
Constant	2.08***	4,747.77***	1.86***	4,764.63***
	(0.05)	(5.81)	(0.08)	(6.89)
Fixed Effects Regression	-	-	Yes	Yes
Player Dummy Variables	Yes	Yes	-	-
Observations	2,240	2,240	2,240	2,240
R-squared	0.14	0.96	0.16	0.19
Number of Players	200	200	201	200

Robust standard errors in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The shorthand names of various variables were used above. *TmStrengthDif* is simply the Team Strength Difference variable. *PerformanceDev3* is the players season average subtracted from his average performance in the last 3 games. *PlayedYesterday* and *OppPlayedYesterday* indicate whether the player played yesterday and his opponent played yesterday, respectively.

Table 3: Pooled Regressions With Various Specifications

Variables	(1) Fantasy Ponts	(2) Price	(3) Fantasy Points	(4) Price
HomeIce	0.32***	-11.71	0.39***	-13.49
	(0.12)	(22.80)	(0.11)	(10.40)
SeasonAverage	0.76***	2,005.62***		
	(0.09)	(20.87)		
TmStrengthDif	-0.25	19.98	0.98**	13.03
	(0.35)	(69.87)	(0.49)	(44.83)
PerformanceDev3	0.03	26.51***		
	(0.04)	(7.61)		
PerformanceDev5			-0.28***	151.47***
			(0.06)	(5.92)
PlayedYesterday	-0.15	1.85	-0.16	-4.79
	(0.15)	(30.46)	(0.15)	(14.64)
OppPlayedYesterday	0.08	-9.93		
	(0.16)	(29.42)		
Constant	0.14	38.00	1.84***	4,774.34***
	(0.22)	(48.75)	(0.08)	(8.21)
Player Dummy Variables	-	-	Yes	Yes
Observations	2,240	2,240	2,240	2,240
R-squared	0.04	0.87	0.16	0.98

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

The shorthand names of various variables were used above. *TmStrengthDif* is simply the Team Strength Difference variable. *PerformanceDev3* is the players season average subtracted from his average performance in the last 3 games. *PerformanceDev5* is the players season average subtracted from his average performance in the last 5 games. *PlayedYesterday* and *OppPlayedYesterday* indicate whether the player played yesterday and his opponent played yesterday, respectively.

Table 4: Fixed Effects Regressions

With Controls For First Degree Auto Regressive Disturbance

Variables	(1) Fantasy Points	(2) Price	(3) Fantasy Points	(4) Price
HomeIce	0.12	4.26	0.28**	-1.76
	(0.13)	(5.37)	(0.12)	(5.03)
TmStrengthDif	0.61	7.83	0.93*	4.00
	(0.47)	(19.67)	(0.48)	(19.72)
PerformanceDev3	-1.25***	55.91***		
	(0.06)	(2.53)		
PerformanceDev5			-1.88***	93.82***
			(0.10)	(4.42)
PlayedYesterday	-0.41***	4.06		
	(0.12)	(4.87)		
OppPlayedYesterday	-0.07	-7.70		
	(0.16)	(6.88)		
Constant	2.85***	8,006.28***	2.38***	7,920.85***
	(0.05)	(2.18)	(0.05)	(2.17)
Observations	2,040	2,040	2,040	2,040
R-squared	0.20	0.21	0.17	0.20
Number of Players	200	200	200	200

Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

The shorthand names of various variables were used above. *TmStrengthDif* is simply the Team Strength Difference variable. *PerformanceDev3* is the players season average subtracted from his average performance in the last 3 games. *PerformanceDev5* is the players season average subtracted from his average performance in the last 5 games. *PlayedYesterday* and *OppPlayedYesterday* indicate whether the player played yesterday and his opponent played yesterday, respectively.

Table 5: Random Effects Regressions

	(1)	(2)
Variables	Fantasy Points	Price
HomeIce	0.32**	-2.28
	(0.13)	(12.60)
TmStrengthDif	0.20	19.88
	(0.38)	(50.42)
PerformanceDev3	0.01	85.25***
	(0.04)	(6.33)
PlayedYesterday	-0.14	12.01
	(0.15)	(12.70)
OppPlayedYesterday	0.05	-40.82***
	(0.17)	(15.39)
Constant	1.92***	4,804.35***
	(0.09)	(103.02)
Observations	2,240	2,240
Number of Players	200	200
R-squared	0.01	0.19

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Team Strength and Recent Performance Variable Substitutions

Variables	(1) Fantasy Points	(2) Price	(3) Fantasy Points	(4) Price
HomeIce	0.38***	-1.65	0.38***	-1.58
	(0.14)	(12.36)	(0.14)	(10.13)
OppWinningPercentage	-0.47	49.64	-0.47	43.99
	(0.50)	(46.91)	(0.50)	(43.70)
Last3GamesAveragePoints	-0.18***	89.97***		
	(0.04)	(6.20)		
Last5GamesAveragePoints			-0.29***	153.72***
			(0.05)	(9.34)
PlayedYesterday	-0.18	12.16	-0.16	2.05
	(0.15)	(12.65)	(0.15)	(11.81)
OppPlayedYesterday	-0.04	-39.18**	-0.02	-47.40***
	(0.18)	(15.08)	(0.18)	(13.39)
Constant	2.56***	4,528.74***	2.79***	4,391.90***
	(0.29)	(24.60)	(0.30)	(26.65)
Fixed Effects Regression	Yes	Yes	Yes	Yes
Observations	2,240	2,240	2,240	2,240
R-squared	0.02	0.23	0.02	0.36
Number of Players	200	200	200	200

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Team Strength and Recent Performance Variable Substitutions

With Pooled Regressions

Variables	(1) Fantasy Points	(2) Price	(3) Fantasy Points	(4) Price
SeasonAverage	0.73***	1,979.86***	0.70***	1,945.95***
	(0.10)	(23.49)	(0.10)	(25.11)
HomeIce	0.33***	-10.92	0.33***	-9.44
	(0.11)	(22.78)	(0.11)	(22.59)
OppWinningPercentage	-0.59	-86.19	-0.58	-80.07
	(0.49)	(98.39)	(0.49)	(97.80)
Last3GamesAveragePoints	0.03	26.34***		
	(0.04)	(7.60)		
Last5GamesAveragePoints			0.06	55.57***
			(0.05)	(10.08)
PlayedYesterday	-0.15	2.11	-0.15	-0.83
	(0.15)	(30.50)	(0.15)	(30.53)
OppPlayedYesterday	0.08	-10.38	0.08	-12.34
	(0.16)	(29.42)	(0.16)	(29.43)
Constant	0.43	79.45	0.44	90.99
	(0.33)	(66.62)	(0.33)	(66.48)
Observations	2,240	2,240	2,240	2,240
R-squared	0.04	0.87	0.04	0.87

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Team Winning Percentages
Throughout 2013-2014 NHL Season

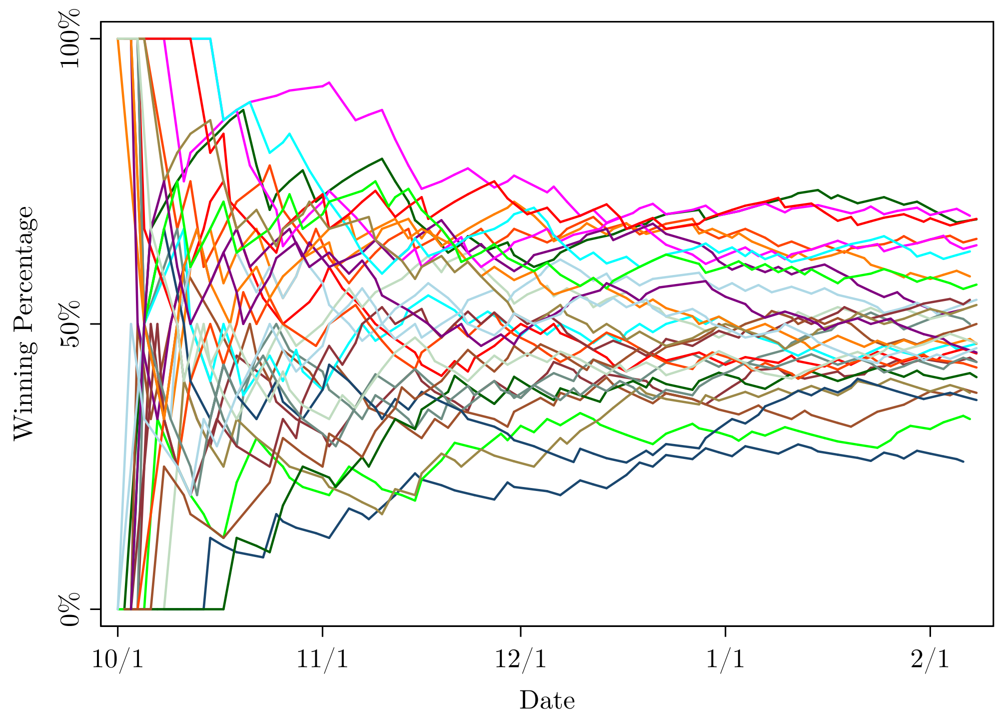

Figure 2: Kernel Density Plots of Season Average and Daily Performance

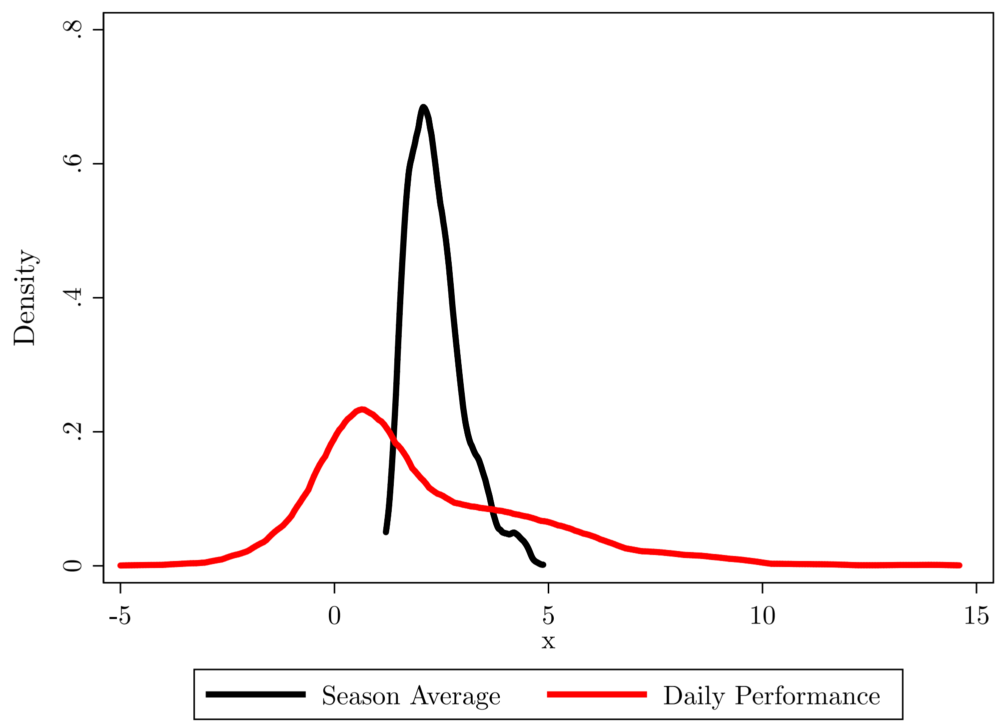

Figure 3: Residuals Against Fitted Values
Main Fixed Effects Estimation Equation