A Statistical Analysis of American Football Running Backs’ Draft Position and Subsequent Earnings

By

Nathan Eckstein

Presented to the Department of Economics, University of Oregon, in partial fulfillment of the requirements for honors in Economics

Prepared under the supervision of
Professor Trudy Ann Cameron

June 2009
A Statistical Analysis of American Football Running Backs’ Draft Position and Subsequent Earnings

By

Nathan Eckstein

Abstract:

This paper describes an empirical analysis of draft position and subsequent salaries as a function of the NFL combine statistics, among other measurable qualities, for running backs from the draft classes of 2005-2008. The findings show that among the top 50 draft picks, a change in draft pick has a much greater effect on income than it does for running backs drafted after the 50th pick. Regression analysis demonstrates that speed, measured at the NFL combine by the times for the 10-yard, 20-yard, and 40-yard dash relative to the player’s weight appears to be the most important quality NFL teams look for in running backs. This analysis should provide insight on just how important combine performances are for incoming professional running backs, and which events they should exert the most energy preparing for.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>Data</td>
<td>6</td>
</tr>
<tr>
<td>Model</td>
<td>12</td>
</tr>
<tr>
<td>Top 50</td>
<td>14</td>
</tr>
<tr>
<td>Average Speed Model</td>
<td>18</td>
</tr>
<tr>
<td>Forecasting</td>
<td>21</td>
</tr>
<tr>
<td>Game Theory Scenario</td>
<td>23</td>
</tr>
<tr>
<td>Caveats</td>
<td>28</td>
</tr>
<tr>
<td>Conclusion</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>36</td>
</tr>
</tbody>
</table>
Introduction:

The NFL combine is now a nationally televised event in which the top prospects from college football, who are heading to the NFL draft, come together to be assessed based on dozens of physical and cognitive tests. Everyone from diehard fans to NFL team scouts watch carefully as potential football stars of tomorrow show off their staggering speed, quickness, strength, agility, jumping ability, and IQ. The tests performed at the combine include the 225-lb bench press, standing vertical jump, broad jump, three-cone drill, 20-yard shuttle, 60-yard shuttle, position drills, measurements, NFL team interviews, the Wonderlic test (IQ test), injury evaluations, the Cybex Machine test (to test the agility and strength of the knees), drug test, 10-yard dash, 20-yard dash, and the glamour event of the combine, the 40-yard dash.

Intuition suggests that the 40-yard dash, relative to the player's weight, would have an important effect on how early a running back gets drafted in the NFL. The reason the 40-yard dash is monitored so closely for running backs is because it is the best measure of top-end speed. The 10-yard and 20-yard dashes give the players less opportunity to hit top speeds, but these shorter sprints tend to measure quickness in the initial take-off very well. Alternatively, the 40-yard dash evaluates player’s quick take-offs and gauges their high top speeds by giving them enough time to reach their full stride.

Although running backs need both quickness and speed, the 40-yard dash is very informative. For example, if the running back breaks through the preliminary line of scrimmage and into the back field, the first forty yards are going be what decides whether he outruns everyone and scores a touchdown, or gets tackled for just a medium gain in yardage. For this reason, the models in this paper devote particular attention to the 40-yard dash, especially when combined with a size variable. This measure of speed can be expected to be the most important
determinant of how soon a running back gets drafted, and therefore how much he is paid, in the NFL.

Running backs train intensely in the hopes that their NFL combine performance, along with other attributes, will cause an NFL team to use one of their team's allotted draft picks to select them. Many players even hire personal trainers to trim unwanted fractions of seconds off their 40-yard dash time and help them improve their performance on the other tests given at the NFL combine. The TEST Sports Football Academy in New Jersey has become a very popular training facility for players preparing for the NFL combine. It offers a 6-8 week program including personal trainer, dieticians, high-tech state-of-the-art facilities, weight rooms, a full turf field, and room and board for all its members. Its goal is to provide players with training that will make them look good in front of the NFL scouts at the combine. To be drafted represents an opportunity to set up a long and profitable NFL career. Many players cannot afford to attend a camp such as this however, and must train on their own, and will have to hope for the best.

The NFL draft is set up so that the NFL team with the worst record the previous season gets the first draft pick, which they can use to select any incoming player who has declared for the draft. The team with the best record from the previous season gets the last draft pick in the first round and has to select the best remaining player that has not already been chosen in the draft by another team ahead of them. The NFL uses this algorithm to allocate new players in the hopes that it will help level the playing field for upcoming seasons and produce the most evenly competitive league possible. There are now only seven rounds in the NFL draft, although there used to be many more. Incoming players want to be selected as early as possible in the draft,

Footnote: Many NFL players do not get drafted. These “Undrafted Free-agents” and can still try-out for NFL teams and if they perform well they can still sign NFL contracts.
because that means they are recognized as the best available recruit not previously selected. On average, the earlier a player is selected, the more their initial NFL contract is worth.

The objective of this paper is to consider running backs coming out of college and determine which attributes NFL teams are looking for. Is it speed, strength, jumping ability, IQ, physical durability, or is it something less measurable like leadership, work ethic or 'heart'? Once the statistically most relevant factors are identified, this empirical model's results will recommend which attributes running backs should exert the most energy to optimize. It may also be able to predict who will be drafted when, and how much they will get paid.

This paper does not delve into what happens after a running back signs his initial contract. It cannot link these attributes used in this paper to players' performances in the NFL. This is important because players may be drafted late and sign initial NFL contracts that pay very little. These same players often end up surprising people and renegotiating for more money later in their careers. This paper does not account for unseen potential in the drafting process and does not look far enough into the future to see whether or not the draft pick turned out to be a good player over the long run.

Data:

Data were collected for 62 running backs from the four annual draft classes spanning the years 2005-2008. These data were chosen because they include measures of running backs' skills and potential before they enter the NFL draft. The constructed data set includes three major categories of variables.

(a) College Rushing Statistics: This category includes rushing yards (YDS), touchdowns (TDS), carries (CARRY- or attempts), yards per carry (YDPERC- how many yards on average the running back gains every time he is handed the ball), and touchdowns per carry (TDPERC).
None of these variables except YDPERC end up having statistically significant coefficients when used to explain draft and income, but empirical analysis is necessary to demonstrate this lack of influence. These variables are still important to consider, because to truly understand the factors which contribute to draft position, we need to know not only what is significant, but also what is not.

(b) NFL Combine Statistics: The NFL combine examines players who are in the off-season after their college careers, but before the NFL draft. The data set only includes the tests given at the combine that are most relevant to running backs, because many running backs do not participate in every test offered at the combine. The variables created from the statistics at the combine are WEIGHT, HEIGHT, times for the 10-yard dash (TEN), 20-yard dash (TWENTY), and 40-yard dash (FORTY), vertical jump (VERT- the player’s ability to jump vertically from a standing position with one step to create momentum), and standing broad jump (BROAD- the players ability to jump forward from a still position).

(c) Draft Selection and Income Statistics: This includes information for the variables DRAFT, INCOME, FIRST50 and YEAR. DRAFT was the player’s overall selection position in the NFL draft. Important to recognize that DRAFT and INCOME are strongly inversely related. FIRST50 is a dummy variable that takes on a value of 1 if the player was selected as a top 50 draft pick, and 0 if he was selected after that (selected between 50 and 250). YEAR is

Foot Note: The College Rushing Statistics are from “Reference 1”. The NFL Combine Statistic information is from “Reference 2” and a more in-depth description of all the tests given and processes of measurement of these tests can be seen at “Reference 3”. The information that provided the data on who got drafted and when is from “Reference 4”. All the players who were drafted and participated in the combine have their initial salary contracts for the NFL visible on “Reference 5”. The information on NFL contracts is shown at “Reference 6”. 
the year in which the player was drafted. INCOME was defined as the per-year salary from the player’s initial NFL contract with the team that drafted him, although players receive many other benefits.

The variable INCOME does not give a complete description of earnings of NFL players without knowing more about the way they are paid and their additional benefits. NFL players get paid every other Friday during the regular season only (9 pay days a year). If injured during football related activities players continue to get paid in full. NFL contracts often included signing bonuses, roster bonuses (for playing a certain amount of games), work out bonuses (for working out with the team a certain amount of days), and incentive bonuses (for reaching certain goals statistically or sometimes based on playing time). When traveling to away games, contracts also cover airfare, hotels and three meals a day. Yet, contracts give the NFL the right to use the players name and picture for any NFL marketing and void the player’s right to take part in dangerous activities or activities that reflect a bad image on the NFL. For this paper only base salary (including signing bonus) is recorded for the variable INCOME.

Table 1 gives a description of all the basic variables that are available for analysis in this paper. The two different dependent variables in this paper, INCOME and DRAFT are shown at the top of Table 1. They are followed by the most significant independent variables, which are then followed by the least significant variables. The variables not included in Table 1 that have not already been mentioned are the interaction terms that combine pairs of the previously mentioned variables into one.

FORTY_WEIGHT is an interaction term between the player’s 40-yard dash times and the
Table 1: List of All Relevant Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCOME</td>
<td>62</td>
<td>1.539</td>
<td>2.155</td>
<td>0.3100</td>
<td>10.00</td>
<td>Million$ / year</td>
</tr>
<tr>
<td>DRAFT</td>
<td>62</td>
<td>105.9</td>
<td>77.80</td>
<td>2.000</td>
<td>250.0</td>
<td>Selection</td>
</tr>
<tr>
<td>FORTY</td>
<td>62</td>
<td>454.9</td>
<td>13.21</td>
<td>424.0</td>
<td>486.0</td>
<td>.01 seconds</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>62</td>
<td>217.59</td>
<td>1.535</td>
<td>18.70</td>
<td>26.70</td>
<td>Lbs</td>
</tr>
<tr>
<td>YDPERC</td>
<td>62</td>
<td>5.405</td>
<td>0.8745</td>
<td>3.368</td>
<td>7.720</td>
<td>Yards/carry</td>
</tr>
<tr>
<td>TEN</td>
<td>62</td>
<td>156.7</td>
<td>5.632</td>
<td>140.0</td>
<td>1.690</td>
<td>.01 seconds</td>
</tr>
<tr>
<td>TWENTY</td>
<td>62</td>
<td>263.3</td>
<td>7.921</td>
<td>241.0</td>
<td>283.0</td>
<td>.01 second</td>
</tr>
<tr>
<td>HEIGHT</td>
<td>62</td>
<td>70.87</td>
<td>1.894</td>
<td>66.00</td>
<td>76.00</td>
<td>Inches</td>
</tr>
<tr>
<td>BROAD</td>
<td>62</td>
<td>118.6</td>
<td>6.461</td>
<td>103.0</td>
<td>130.0</td>
<td>Inches</td>
</tr>
<tr>
<td>VERT</td>
<td>62</td>
<td>34.03</td>
<td>2.614</td>
<td>27.50</td>
<td>40.50</td>
<td>Inches</td>
</tr>
<tr>
<td>YDS</td>
<td>62</td>
<td>2387</td>
<td>1100</td>
<td>510.0</td>
<td>5040</td>
<td>Yards</td>
</tr>
<tr>
<td>TDS</td>
<td>62</td>
<td>23.20</td>
<td>12.81</td>
<td>7.000</td>
<td>62.00</td>
<td>Touchdowns</td>
</tr>
<tr>
<td>CARRY</td>
<td>62</td>
<td>442.1</td>
<td>200.0</td>
<td>136.0</td>
<td>1015</td>
<td>Carries</td>
</tr>
<tr>
<td>FIRST50</td>
<td>62</td>
<td>0.2903</td>
<td>0.4576</td>
<td>0.000</td>
<td>1.000</td>
<td>Yes / No</td>
</tr>
<tr>
<td>YEAR</td>
<td>62</td>
<td>2006</td>
<td>1.096</td>
<td>2005</td>
<td>2008</td>
<td>Years</td>
</tr>
</tbody>
</table>

Note: Dependent variables names are italicized at top of table
player's weight. The conjecture in this paper is that the most important factor determining how early an NFL running back will be drafted is speed and size. Out of the variables collected in the data set, perhaps the FORTY is the best overall measure of speed, and WEIGHT is the best measure of size. The reason why HEIGHT may not be as important a factor when defining size is that the heights of running backs involve very little deviation from the mean of 70.87 inches. The minimum height in the sample is 66 inches, and the maximum is 76 inches, but the distribution has a standard deviation of only 1.894. When tackling someone weight plays a much larger role than height because weight affects momentum, which is what the opposing defensive player is combating.

To look at the influence of the interaction term FORTY_WEIGHT we can analyze the derivatives of the model when FORTY and WEIGHT are also included in the regression.

If \( \text{INCOME} = C + b_1 \cdot \text{FORTY} + b_2 \cdot \text{WEIGHT} + b_3 \cdot \text{FORTY}_\text{WEIGHT} + \ldots \)
then \( \frac{d\text{INCOME}}{d\text{FORTY}} = b_1 + b_3 \cdot \text{WEIGHT} \)

And \( \frac{d\text{INCOME}}{d\text{WEIGHT}} = b_2 + b_3 \cdot \text{FORTY} \)

Without the new variable the derivatives would be constant, and equal to \( b_1 \) and \( b_2 \) respectively. This higher order variable allows WEIGHT to have an effect on the derivative of INCOME with respect to FORTY and vice versa, which can give the model a better fit and increase the adjusted R-squared.

Foot Note: The above section focuses on FORTY_WEIGHT because it was the only statistically significant of the three interaction term variables that were generated.
Having either a lightning fast running back or a bruising, strong running back can prove to be a huge asset for NFL teams. This being said, smaller backs tend to be the faster backs and vice versa. It is much more impressive for a big running back (such as Brandon Jacobs, who is 6'4” and 267 lbs) to be able to run a 4.6 second 40-yard dash, than for a small running back (such as Wali Lundy who is 5'11’ and 214 lbs) to do the same. This phenomenon can be better measured by the variable FORTY_WEIGHT instead of just including both FORTY and WEIGHT in the regression individually. This coefficient will read whether it is an asset for running backs entering the NFL draft to not simply be big or fast, but to be a hybrid player like Brandon Jacobs who exhibits both attributes. Essentially it measures the running back’s ability to create momentum and fall forward, which is critical in a sport where every inch matters.

Without this statistical adjustment then two running backs such as Brandon Jacobs and Wali Lundy would be seen as equals in terms of their slopes with respect to speed. That would be inaccurate, as we can see that Brandon Jacobs, the bigger running back, got drafted 60 spots earlier and gets paid $927,000 more a year for running the same 40-yard dash time. When these two athletes are ranked with this FORTY_WEIGHT variable alone Jacobs ranks 2\textsuperscript{nd}, while Lundy ranks 48\textsuperscript{th}.

Another interaction term that helps clarify this situation is FORTY2. FORTY2 or forty-squared is equal to the player’s 40-yard dash time multiplied by itself, or squared. This variable shows the effects on a player’s income of having incrementally more speed. In a sense it amplifies the original speed variable, FORTY, to give a more in depth look at the effects of having that little bit of extra speed. With all four of these variables, FORTY, FORTY2,
FORTY_WEIGHT, and WEIGHT, included in a regression the analysis of speed relative to size pertaining to the income of running backs is more complete.

Models:

There are two candidate dependent variables for this paper. These two variables are DRAFT, which measures the outcome of when the running back was selected overall in the NFL draft in a given year, and INCOME, which is measured by the salary contract the running back signed with the NFL team that drafted him (in units of millions of dollars per year). The two variables are strongly inversely correlated because the better the player, the earlier he will be drafted (represented by a small number in the data) and the more he will get paid in his initial NFL contract (represented by a large number in the data).

The initial model involves simply putting all of the independent variables from both the player’s combine performance and his college career into the specification and seeing how well they explain a running back’s draft position. Estimates for this model are shown in Table 2. All of the coefficients of the explanatory variables except FORTY are statistically insignificantly different from 0 because their t-statistics are all below the absolute value of 1.96. Another weakness of this first model is that the adjusted R-squared value is only 0.4191, meaning that over half of the variation in the dependent variable DRAFT is not explained by the independent variables in this initial model.

To improve upon the first model, INCOME was used as an alternative dependent variable instead of DRAFT. In this model, FORTY is more significant than in the initial model. The next step is to drop the individually explanatory insignificant variables and to add higher-order terms associated with the one significant variable, FORTY, such as the square of FORTY, denoted
### Table 2: Initial Draft Model (OLS, n=62)

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEN</td>
<td>-67.47</td>
<td>269.8</td>
<td>-0.250</td>
<td>0.8035</td>
</tr>
<tr>
<td>TWENTY</td>
<td>-169.8</td>
<td>323.0</td>
<td>-0.5255</td>
<td>0.6015</td>
</tr>
<tr>
<td>FORTY</td>
<td>410.2</td>
<td>165.6</td>
<td>2.476**</td>
<td>0.0166</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>-0.3068</td>
<td>0.8078</td>
<td>-0.3798</td>
<td>0.7056</td>
</tr>
<tr>
<td>HEIGHT</td>
<td>-2.691</td>
<td>5.573</td>
<td>-0.4828</td>
<td>0.6313</td>
</tr>
<tr>
<td>BROAD</td>
<td>-2.202</td>
<td>2.002</td>
<td>-1.099</td>
<td>0.2767</td>
</tr>
<tr>
<td>VERT</td>
<td>-3.217</td>
<td>4.116</td>
<td>-0.7816</td>
<td>0.4381</td>
</tr>
<tr>
<td>YDS</td>
<td>0.009475</td>
<td>0.01524</td>
<td>0.6214</td>
<td>0.5371</td>
</tr>
<tr>
<td>TDS</td>
<td>-0.6814</td>
<td>1.153</td>
<td>-0.5905</td>
<td>0.5574</td>
</tr>
<tr>
<td>YDPERC</td>
<td>-12.13</td>
<td>10.17</td>
<td>-1.192</td>
<td>0.2386</td>
</tr>
<tr>
<td>C</td>
<td>-520.3</td>
<td>615.0</td>
<td>-0.8459</td>
<td>0.4015</td>
</tr>
</tbody>
</table>

** Statistically significant at the 5% level
R-squared: 0.5143
Adjusted R-squared: 0.4191
Log likelihood: -335.0

### Table 3: Final Model- INCOME as a function of selected factors (OLS, n=62)

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAFT</td>
<td>-0.009781</td>
<td>0.003520</td>
<td>-2.778**</td>
<td>0.0075</td>
</tr>
<tr>
<td>FORTY</td>
<td>-3.714</td>
<td>0.9688</td>
<td>-3.833**</td>
<td>0.0003</td>
</tr>
<tr>
<td>FORTY2</td>
<td>4.948</td>
<td>12.85</td>
<td>3.850**</td>
<td>0.0003</td>
</tr>
<tr>
<td>FORTY WEIGHT</td>
<td>-3.835</td>
<td>1.497</td>
<td>-2.562**</td>
<td>0.0132</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>1.768</td>
<td>6.840</td>
<td>2.585**</td>
<td>0.0124</td>
</tr>
<tr>
<td>YDPERC</td>
<td>0.5898</td>
<td>0.2506</td>
<td>2.353**</td>
<td>0.0222</td>
</tr>
<tr>
<td>C</td>
<td>659.3</td>
<td>193.2</td>
<td>3.411</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

** Statistically significant at the 5% level
R-squared: 0.5240
Adjusted R-squared: 0.4721
Log likelihood: -112.1

Foot Note: *EVEiWS 6 is the software package used to produce the estimates reported in this paper.*
FORTY^2 and FORTY_WEIGHT. This regression is shown in Table 3. It features six different statistically significant variables including draft, which is now being used as an important explanatory variable for INCOME. Notice also that Table 3 has a higher Adjusted R- squared than Table 2, although these statistics are not directly comparable because the dependent variables are different.

The final model features only three of the original variables that describe players measured physical attributes in the data set (FORTY, YDPERC, and WEIGHT). The other three variables are DRAFT, and higher order terms FORTY^2 and FORTY_WEIGHT. Intuitively it makes sense that FORTY would have a negative coefficient because when running the 40-yard dash the best players have the fastest times (represented by a small number) and the biggest salaries (represented by a large number). This negative relationship is not seen in WEIGHT or YDPERC meaning that heavier players get paid more, as do players who have higher average yards per carry in college. Also notice that the sign on the coefficient for draft is negative meaning that the earlier a running back is drafted the higher will be his income when he signs his initial NFL contract.

**Top 50:**

One thing that is instantly clear when looking at the graph of INCOME, plotted against DRAFT is that there are very high rewards for being an early draft pick. Teams are willing to offer the really exceptional running backs coming into the NFL a lot of money to sign with their team. After the first group of elite running backs are selected in the draft, all of the remaining running backs get paid somewhat similarly in their initial NFL contracts. In fact, you could argue that the data should be split into two different data sets because income is so much higher for the top-echelon running backs than for the rest of the group.
Figure 1 shows a graph of the raw data for INCOME plotted against DRAFT position. Figure 1, also includes a linear prediction of a curved best fit line that was generated to help predict INCOME based on DRAFT. It is quickly noticeable that the players drafted about 50th or later in the overall selection are very similar when it comes to income. Players drafted earlier than 50th have a wide variety of incomes that range much higher than the later draft picks. For this reason the dummy variable First50 can be included in regressions to explain the difference in Figure 2, which is a graph of the linear prediction of the logarithm of INCOME on DRAFT, with the later draft picks range from around $300,000 to $700,000 a year. A typical NFL contract for lower level players is $530,000 a year (this is the median of the later-drafted player’s salaries).

Table 4 features a regression in which the log of income is solely described by variables related to draft position in the data set. The regression includes DRAFT, FIRST50, YEAR, and an interaction term called DRAFT_FIRST50 (a generated variable equal to Draft multiplied by the dummy variable First50). All four of the independent variables prove to be statistically significant, having t-statistics with absolute values greater than 1.96. As expected the variable First50 was overwhelmingly the most significant with a coefficient of 2.327 and a t-statistic of 18.76.

\[
\frac{\Delta INCOME}{\Delta FIRST50} = 2.327 + (-0.04211) \times DRAFT
\]

So the premium for being a top 50 running back initially increases INCOME by very large amount of $2,327,000, but also declines by $42,000 for each position in the draft due to the interaction term’s, DRAFT_FIRST50, coefficient. To intuitively understand the mathematics we can plug in two numbers for the variable DRAFT and see how it changes the equation. We can see that getting selected 30th in the draft vs. getting selected 40th in the draft changes INCOME (millions of dollars per-year) by very different amounts.
Figure 1: Raw data on INCOME vs. DRAFT position with curved linear prediction line

Note: Notice the most dramatic part of the curve, and change in slope, takes place close to the 50th draft pick. This model assumes that income varies inversely with draft position. The R-squared value is over .78 in this specification.

Figure 2: Graph of the linear prediction for a model of LGINC regressed on draft position

Note: Separating the slope of the top 50 draft picks from the rest, showing that being drafted one spot higher when in the top 50 represents a greater gain in income than a one spot change for players drafted below 50.
### Table 4: Income/Draft Model- LGINC as a function of draft statistics (OLS, n=62)

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAFT</td>
<td>-0.001750</td>
<td>0.000553</td>
<td>-3.168**</td>
<td>0.0025</td>
</tr>
<tr>
<td>DRAFT_FIRST50</td>
<td>-0.04211</td>
<td>0.003459</td>
<td>-12.17**</td>
<td>0.0000</td>
</tr>
<tr>
<td>FIRST50</td>
<td>2.327</td>
<td>0.1240</td>
<td>18.76**</td>
<td>0.0000</td>
</tr>
<tr>
<td>YEAR</td>
<td>-0.05964</td>
<td>0.02780</td>
<td>-2.144**</td>
<td>0.0362</td>
</tr>
<tr>
<td>C</td>
<td>119.3</td>
<td>55.79</td>
<td>2.139</td>
<td>0.0367</td>
</tr>
</tbody>
</table>

** Statistically significant at the 5% level

R-squared: 0.9354
Adjusted R-squared: 0.9309
Log likelihood: 4.834
\[
\frac{\Delta \text{INCOME}}{\Delta \text{FIRST50}} = 2.327 + (-0.04211) \times 40 = .6426
\]

\[
\frac{\Delta \text{INCOME}}{\Delta \text{FIRST50}} = 2.327 + (-0.04211) \times 30 = 1.064
\]

The variable for YEAR unexpectedly has a negative coefficient, meaning running backs that were drafted in earlier years got paid more. This could be the result of changing demand for running backs between different year’s drafts. Of the 5 running backs selected as top 5 overall draft picks between 2005-2008, 4 of them were from the first 2 year’s draft classes (2005-2006) and 3 were from 2005. The 2005 draft class that was so well endowed with running back talent that it could explain why the variable YEAR has a negative coefficient, when one would intuitively expect positive inflation to cause a positive coefficient. If the data spanned over more years this variable would be more relevant and probably have a positive coefficient.

The R-squared of this model is 0.9354 which is much larger than the R-squared of 0.5240 of the final model which featured the 40-yard dash, weight and yard per carry as the major determinates of income. This makes sense because it is expected that the earlier draft picks will be paid more. In fact, it could be the case that FORTY, WEIGHT and YDPERC determine DRAFT and when DRAFT is used is this new model, it merely serves as a substitute for these variables. It is important to remember that the two R-squared statistics are not directly comparable because the models have to different dependent variables; it is just looked at to analyze the overall fit of models.

**Average Speed Model:**

One issue that must be addressed in this paper is that there is some multicollinearity between the three variables that measure the times for the 10-yard, 20-yard, and 40-yard dashes. All three variables are slightly different measures of the running back’s sprinting ability, so it makes sense that they would be highly correlated. These high correlations are given in Table 5.
To solve this problem without completely dropping any of these speed variables from the regression, calculations were done to convert all of the information held in the 10-yard, 20-yard, and 40-yard dashes into a variable that measures the average miles per hour during all three of these tests. This was done by some basic conversion of units and then averaging the three remaining numbers into one.

Note:

- 1 mile = 1760 yards
- 1 hour = 3600 seconds

\[
\frac{\text{length of dash "yards"}}{\text{dash time "seconds"}} \times \frac{1 \text{ mile}}{1760 \text{ yards}} \times \frac{3600 \text{ seconds}}{1 \text{ hour}} \quad \text{Leaving units in MPH, So...}
\]

\[
\frac{3600 \text{ seconds}}{1760 \text{ yards}} \quad \text{Reduces to} \quad \frac{45 \text{ seconds}}{22 \text{ yards}} = \text{Conversion factor from yards per second to MPH}
\]

\[
\frac{\text{seconds}}{\text{yard}} \times \frac{(45 \text{ seconds})}{(22)} = \frac{\text{miles}}{\text{hour}}
\]

So...

\[
\left( \frac{40 \text{ forty yard dash time}}{22 \text{ yards}} + \frac{20 \text{ twenty yard dash time}}{22 \text{ yards}} + \frac{10 \text{ ten yard dash time}}{22 \text{ yards}} \right) \times \frac{45 \text{ seconds}}{22 \text{ yards}}
\]

= average speed during the three sprints in MPH

With this variable created for the running back’s three different sprint times we can see the effect of all the sprints performed by the running backs at the NFL combine and not just the 40-yard dash. Similarly to how the FORTY was used in the final model, this new variable AVESPEED can be used to explain INCOME, not only through its linear form, but also via a quadratic form (AVESPEED^2), and via the interaction term AVESPEED_WEIGHT. Looking at Table 6, one will notice that all three of these AVESPEED terms are statistically significant when used to explain INCOME, with DRAFT, WEIGHT and YDPERC also included as
Table 5: Correlations of Ten, Twenty and Forty

Included observations: 62

<table>
<thead>
<tr>
<th></th>
<th>Ten</th>
<th>Twenty</th>
<th>Forty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ten</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twenty</td>
<td>0.8473**</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Forty</td>
<td>0.07690</td>
<td>0.9195**</td>
<td>1.000</td>
</tr>
</tbody>
</table>

** Covariance greater than .800 (could feature multicollinearity problems)

Note: This high correlation shows multicollinearity between the three variables. Some economists suggest that if the correlation is .800 or above, that one or more variables should be taken out of the model or converted into another variable that will not have this problem. The two cases of covariance’s greater than .800 will not be a issue in Table 6 because all three of the variables were combined into one variable for “average speed” measured in miles per hour.

Table 6: INCOME as a function of AVESPEED (OLS, n=62)

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAFT</td>
<td>-0.0128</td>
<td>0.003594</td>
<td>-3.582**</td>
<td>0.0007</td>
</tr>
<tr>
<td>AVESPEED</td>
<td>-98.29</td>
<td>40.91</td>
<td>-2.402**</td>
<td>0.0197</td>
</tr>
<tr>
<td>AVESPEED2</td>
<td>2.5633</td>
<td>1.075</td>
<td>2.382**</td>
<td>0.0207</td>
</tr>
<tr>
<td>AVESPEED_WEIGHT</td>
<td>0.08843</td>
<td>0.04291</td>
<td>2.060**</td>
<td>0.0441</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>-1.349</td>
<td>0.6624</td>
<td>-2.037**</td>
<td>0.0464</td>
</tr>
<tr>
<td>YDPERC</td>
<td>0.6129</td>
<td>0.2689</td>
<td>2.279**</td>
<td>0.0266</td>
</tr>
<tr>
<td>C</td>
<td>902.4</td>
<td>380.4</td>
<td>2.372**</td>
<td>0.0212</td>
</tr>
</tbody>
</table>

** Statistically significant at the 5% level
R-squared 0.4524
Adjusted R-squared 0.3927
Log likelihood -116.4

20
explanatory variables. This model does not explain quite as much of the variation in the
dependent variable as the model in Table 3 did, in that it’s adjusted R-squared is 0.3927 and the
final model’s (Table 3) is 0.4721.

This decrease in R-squared is important to recognize, but this AVESPEED model is still
interesting to analyze because it takes into account quickness in acceleration by including
statistics from the 10-yard and 20-yard dashes at the combine. The vast majority of actual
running plays in the NFL involve distances closer to 10 yards than 40. Sometimes not enough
respect is given to the 10-yard and 20-yard dashes, because the 40 yard plays are the highlight
reel plays seen on SportsCenter, where as the shorter plays are less fun to watch, but still impact
the game significantly. In future studies if 10-yard and 20-yard dash became statistically
significant through the gathering of a higher number of observations, this model could help avoid
multicollinearity and still provide information on how the 10-yard and 20-yard dashes explain
INCOME.

Forecasting:

This data set contains information from the draft years 2005-2008, so one inquiry is,
whether this model can be used to predict what will happen for individual running backs from
other year’s drafts? To predict this, the final model from this paper can be used, by inserting a
player’s personal statistics in for the independent variables and multiplying them by the
appropriate estimated coefficients of the model. This predicted number, we will call INCOME*.

Formula:

\[
\text{INCOME}^* = -0.009781(\text{DRAFT}) - 3.714(\text{FORTY}) + 4.948(\text{FORTY2}) -
3.835(\text{FORTY WEIGHT}) + 1.768(\text{WEIGHT}) + 0.5898(\text{YDPERC}) + 659.3
\]
INCOME* can be compared to the player’s actual INCOME to test the in-sample predictive power of the model used in this paper. If this model is perfectly consistent over the years then the estimated INCOME* will be exactly equal to the actual INCOME and the residual will be 0. Of course, depending on the year of the draft some of the statistics may need to be manipulated to account for inflation.

This forecasting procedure will be used to estimate the INCOME of the last three running backs in the NFL that received the league’s MVP award, before 2005. These are the league’s best running backs prior to 2005, the first year that this paper’s data set has information on. So, these MVP winners are exactly what NFL scouts were looking for when they drafted running backs from the sample used in this paper. The three MVPs used in this forecast are Barry Sanders (1997), Terrell Davis (1998), and Marshall Faulk (2000). Before correcting for inflation, the model used in this paper predicted INCOME* equal to $4,364,760 for Barry Sanders, $1,063,440 for Terrell Davis, and $4,728,810 for Marshall Faulk. These estimates are in terms of $ from 2005-2008, and were too high, because in the years these players got drafted the national CPI was lower and NFL players made less money.

To adjust for inflation, the players’ predicted INCOME* was multiplied by a ratio that was equal to the CPI of when they were drafted (April of their draft year), divided by the CPI of the average date when the running backs from this study’s sample were drafted (median between April 05 and April 08 = October 06). The predicted numbers were still too high, even when inflation was considered. This could indicate that athletes now are getting paid more disproportionately to the rest of society than in previous years.

Terrell Davis’s predicted income with inflation, which we will call (INCOME**), from this model was $799,555 a year. He ran a slow 40-yard dash of 4.72 seconds and was drafted
late, but excelled later in his career, and renegotiated for a lot more money. His actual INCOME was $131,000 a year when he entered the NFL. This leaves a residual equal to $668,555.

Marshall Faulk’s INCOME** is $3,447,651 a year. He was drafted second overall in the draft, but his actual INCOME was only $2,457,143 a year. He was still the highest paid rookie that year. The residual for Faulk’s estimate is $990,509.

Barry Sander’s predicted INCOME** was $2,661,228 a year. He was also a top draft pick and when he signed his rookie contract he was the highest paid rookie in NFL history, making $1,220,000 a year. This still leaves Sander’s estimate with a high residual of $1,441,228.

For none of these players does INCOME** come close to predicting reality. In fact, INCOME is only 16% the size of the INCOME** variable for Terrell Davis. For Marshall Faulk INCOME is 71% of the INCOME** variable (which is better), but INCOME is only 46% of the size of INCOME** for Barry Sanders. This could possibly be due to changes in how players are marketed or changes in the amount of money the NFL makes. Also, the farther back in time we try to forecast, the bigger the residual gets. The technological era gives fans more access to these NFL players. Every NFL game is now available on T.V. through a season pass one can buy at NFL Network (satellite T.V. station). Fans can see their favorite players online, on site at the stadiums, and even get complete draft coverage through NFL Network. This among other factors could account for the poor ability of this model to forecast running back’s salaries multiple years back in time.

Scenario:

The empirical information in this paper may be useful to young running backs that are planning a career in professional football. By setting up future expected incomes of NFL players in terms of present discounted net benefits of hiring a personal trainer and making a business like
decision to invest in their football skills prior to the start of their college careers. For this example, imagine a scenario in which an all-state high school running back, with two years left before college, is planning his future. He has already been contacted by football coaches from top universities about signing his letter of intent to play for their school.

This player could use a regression model like this to see that if he could increase his yards per carry throughout his college career by 1, that this improvement appears to increase his per-year NFL salary by $589,800, holding all else constant. This is implied because in the final model, YDPERC has an estimated coefficient of .5898. Assume that the pay increase is for a typical 4-year NFL contract.

Let’s also say that this player has 20 hours per week of free time after school and normal football practice. He believes that if he uses this time to better his running skills by hiring a personal trainer, staying on the field after practice is over and working to improve his; skills, endurance, flexibility, strength and football IQ (film study), he can increase his yards-per-carry in college by one yard, compared to what it would be if he did not put in the extra time. To do this the player would have to hire an experienced, top-of-the-line personal trainer/dietician for $500 a month or $6,000 a year for his two remaining years of high school.

In this example working part time for minimum wage is the players second best alternative or “opportunity cost” of staying after practice and working on his game. So, 20 hours a week multiplied by 50 weeks (2 weeks of vacation a year) = 1000 hours a year multiplied by $6.55 per hour (As of July 24, 2008, the federal minimum wage in the United States) = $6550. Now add this $6,500 in forgone wages to the amount he is paying for the personal trainer to reach an overall opportunity cost of $6,000 + $6,500 = $12,500. This equals the total opportunity cost of extra training for the NFL during high school.
Now that we have the total costs and benefits of the training, we need to calculate the present discounted value of the increase in future expected earnings to determine if it is worth it to follow this plan.

\[ \text{PDV} = \frac{A}{(1+r)^n} \]

Total Costs = $12,500 with n = 1, 2.

Total Benefits = $589,800 with n = 7, 8, 9, 10.

\[ r = 0.05 \text{ (which is to say, 5\% interest) then} \]

\[ \text{PDV} = \frac{12,500}{(1.05)^1} + \frac{12,500}{(1.05)^2} + \frac{589,800}{(1.05)^7} + \frac{589,800}{(1.05)^8} + \frac{589,800}{(1.05)^9} + \frac{589,800}{(1.05)^10} = 802,540 \]

This number is the present discounted value of increasing your yards per carry by 1 yard throughout your college career. Therefore, if in 20 hours a week, for two years, he could increase his yards per carry in college by one yard. Then if the player has the means to pay the personal trainer it would initially seem that, in the long run, his time is better spent working out, lifting weights and studying tape then getting paid to work a part time job after school.

This is not to say that working out takes the same level of effort as working a part time job, but most football players to some extent enjoy the sweat and grind of a good work out. The most significant thing to remember is that in no way is working on your football game a guarantee, injuries do happen. There is also a possibility that the college this person plans to attend has already recruited a running back better than him and he will not be in the spotlight. Then again, maybe the player can’t pass classes and flunks out. All of these potential situations are very real and the player needs to understand this risk before making an informed decision. Almost everything needs to go right for an NFL career to happen. So how can the player take this into account when making his final decision?
Remember we are assuming that this player has already been contacted by major universities about signing his letter of intent so let’s assume he would make it to a college team with or without this extra training. Also this model’s data set only consists of NFL draftees, so we have to assume, either he would make it to NFL without the training and the training just boosts his salary, or that he does not get drafted and the training is a waste of his time and money.

What this model fails to explain is if the training is the difference between getting drafted or not. In that case the payoff would be huge because it is the difference between an NFL salary and a normal job’s salary. Unfortunately, it is beyond the scope of this paper because it would be impossible to estimate accurately what percentage of the time a one yard increase in yards per carry is the difference between being drafted or not. This is due to the fact that the draft process relies on the entirety of the player’s attributes, not just yards per carry. If the scenario could somehow include this third unlikely possibility it would only increase the average probable payout of the investment. Also, this scenario cannot account for the extra benefit associated with the fame gained from football (either in college or the pros) that could have a large impact on employability in other fields.

3.5% of college football players are invited to the NFL scouting combine, the pool from which NFL teams make most of their draft picks. That being said, the player in this example is an all-state high-profile running back not a random college football player. For this reason in this scenario his odds will arbitrarily move up from a normal college football player’s odds of 3.5%

Foot Note: Information on the percentage of high school/college football players that make it into the NFL was gotten off of “Reference 7”
to 20% for being a star athlete.

Notice that even though he is a star athlete he still has an 80% chance of not making it to the NFL. If this quite likely possibility of failure between now and the NFL draft is the case for this player, then this whole process will have been a waste of his time and money. To account for this the present discounted value of training and making it to the NFL should be cut by 80%.

\[
PDV = 802,540 \times 0.20 = 160,508.
\]

And the PDV of investing in the training and not making it to the NFL should be cut by 20%.

\[
PDV = \frac{-12,500}{1 + 0.05} - \frac{12,500}{1 + 0.05^2} = -23,242 \times 0.80 = -18,594
\]

Then add the two numbers to find the average probable payout.

\[
160,508 - 18,594 = 141,914
\]

The common thought process when looking at present discounted values is that if the number is positive, which it is in the case of both the average and all-state players, then the investment should be made. This situation is a little more complex than that because of the extremely high risk of getting a negative payout. In other words this investment plan is only recommended for people who are interested in taking a very high-risk/high-reward investment, similar to a high yield juke bond. Of course, the present discounted value is not the only factor to take into account when this decision is made in real life because many players may be overly confident and believe they have a better chance of making the NFL draft than they actually do. In addition, some people would rather play football and never make a profit, than work part-time.
For them chasing the dream can be rewarding even if the dream never becomes a reality. For these and other reasons we see an inefficiently high proportion of high school players focusing on athletics over school and work preparations.

<table>
<thead>
<tr>
<th></th>
<th>All-State Football Player</th>
<th>Average Football Player</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk</td>
<td>20% chance of positive payout</td>
<td>3.5% chance of positive payout</td>
</tr>
<tr>
<td>Gain if investment succeeds</td>
<td>$802,540</td>
<td>$802,540</td>
</tr>
<tr>
<td>Loss if investment fails</td>
<td>-$23,243</td>
<td>-$23,243</td>
</tr>
<tr>
<td>Average probable payout</td>
<td>$141,914</td>
<td>$5,659</td>
</tr>
</tbody>
</table>

Note: The all-state football player is the one from which this entire scenario has been based. Even though the average probable payout is positive for the training of an average football player, most investors would not recommend this investment strategy because it is so risky. In fact, a lot of investors would not even recommend this plan to the all-state football player because it is not a safe investment and the earnings from the few players that make it to the NFL skews upward the average probable payout for the majority that do not.

Caveats:

This paper was written with a fairly short time constraint due to a change in topic half way through the term. It was a learning experience and can be viewed as a beginning project which may lead to further studies that will take what has been presented here and elaborate on these ideas by analyzing the issues that remain unanswered and learning from the limitations of this paper. One aspect of empirical work is that it is never truly finished. There will always be
ways to further and improve this study. Some ideas for doing this in the future are mentioned in
the following section.

The fairly low adjusted R-squared of the final model suggests that the regression is only
explaining 52.4 percent of the variation of the independent variable. One way in which to
improve this model for elaboration would be to collect more data and increase the sample size
from a low of 62 observations over 4 years, to a value that is more statistically meaningful. With
this increased amount of data the coefficients may show up as more statistically significant and
hopefully the dependent variable can be explained more precisely. Also, if there was a way of
measuring attributes such as personality or leadership during the college career, that information
could possibly be included as additional variables in a regression. For instance, if players were
graded on how personable they were in the interviews given at the NFL combine, that score
could be correlated with draft or income. Another idea is to measure leadership in terms of
whether the player was team captain on his college team as well as how many games or seasons
he spent as team captain.

A further possible determinate that was not measured in this paper which NFL teams
might look at when drafting a running back is the running back’s receiving ability. This is most
often critiqued during his college career and during the running backs’ position drills at the
combine. This information could easily be obtained by gathering player’s college career
receptions, receiving yards and receiving touchdowns. Additional factors could be age, what year
in school the player was when he declared for the draft, whether a player left after only two or
three seasons in college vs. if he prolonged his playing career through graduation.
This would be interesting to look at because on one hand a lot of the really good running backs declare for the draft before graduation because they feel they have nothing left to prove in the college game and are ready for the pros, or they don’t want to risk injury the next year in college and miss out on their opportunity to make a lot of money in the NFL. One the other hand, older players who have declared for the draft are more experienced and often more physically mature then the younger players.

In this data set fullbacks and halfbacks are clumped into one category. This can be misleading because fullbacks are usually drafted for strength and blocking ability, not the speed and running ability, which are so vital to being a good halfback. If all the fullbacks were purged from the data it could give a better look at a more specific type of player than this paper covered.

Occasionally some top athletes don’t participate in the combine because they don’t want to hurt their already high draft stock. For instance, if they already know they are going to be drafted in the top 10 without doing all of the drills at the combine, they may not want to risk a bad performance when they have everything to lose and almost nothing to gain. These players were not included in the sample because of their missing data. If the combine were made mandatory for all incoming draftees, then it would provide more consistent information. Unfortunately, this will probably not be the case for some time as, currently only the good recruits even get invited to the combine.

When looking at ways to improve the estimates from the college performances, one should notice that running for 3,000 career yards in the Pac-10 or any other major conference is very different than doing the same feat at a Division 2 or 3 school, where the competition is much lower. To fix this misleading representation of talent, the use of a dummy variable would
improve the integrity of the numbers. In the dummy variable, running backs would receive a “1” for being in a Division 1 conference and “0” for competing in a Division 2-3 conference. This could help make the yards and touchdown variables portray more accurate talent levels among players. This dummy variable could also be used as an interaction term with other variables, to see if the slopes in the explanatory variables differ across backgrounds.

Another vital aspect that this paper does not cover is that the model has no variables to capture the running back’s winning ability. Recording the player’s total games won in college and/or their winning percentage, even though winning is a team effort, may provide some valuable information on the little things that the player does on the field that aren’t measured statistically, but help the team win. This could be blocking or even attracting defenders to lead them away from a teammate who has the ball when being used as a decoy. This also could be significant because usually the best running backs play for the best teams and therefore might be drafted earlier.

It may be helpful to control for the number of games started in college and how many games were played on national television, or counts of other situations leading to large public exposure (i.e. the Heisman trophy ceremony “top 4 or 5 college players of the year”, All-American game, National Championship, etc.). All of which could provide useful insight into how early the player will get drafted. Even with all the numbers that enter into making a draft selection, holding all else constant, a famous player is going to have a better chance at getting drafted early and paid more than a player who is an unknown.

It might be helpful to look at each running back’s previous year’s running statistics, rather than their career statistics. Use of more recent data could improve the model because it
measures recruiter’s short term memories, instead of their long-term memories. Consider the case of a player who has been pretty good for four years in college, but has shown little improvement over these years vs. a player who was not as good for his first couple of years, but showed improvement and had a standout season his senior year. Measuring career statistics, the way this paper did, would be misleading in this case because usually NFL scouts care more about what kind of player the running back is now, than how long he has been this type of player. Also using the last season’s statistics would make the player’s statistics more comparable because every player would have one season’s worth of yards, touchdowns and carries instead of having varying amounts for those who played more seasons than the others. Potentially the model could feature both career marks and the single season marks previous to the draft, but there could be multicollinearity issues.

There are endless numbers of ideas and information that could be explored, as noted above, that would provide additional explanation on how running backs get drafted. Yet, so much of the game is based on natural instinct and this may not be amenable to any qualitative measurement. It is a feeling, a gut reaction that tells the player when to cut back, when to spin, and the ability to have a keen eye to instantly sum up opponents running at you and make the correct decision on whether to avoid and outrun your opponent, or lower your shoulder and overpower him. These things cannot be measured, and often, cannot be taught. This paper uses empirical data to attempt to explain as much as possible of something that is half science and half human artistry.

While differences in instinct maybe important, the likelihood that there is a substantial amount of latent heterogeneity in the model should not discredit the empirical evidence. As long
as these omitted variables are uncorrelated with the included ones, there will be little bias in the coefficients of the included variables. However, access to these other factors could allow greater precision in estimates. For example, if an NFL team owner is comparing a running back that has a good feeling for the game vs. one that can run a 4.24 second 40-yard dash and has a 10 foot 10 inch broad jump (like Chris Johnson), the owner will likely pick Johnson because of his raw athletic potential. Any well-coached team will find a way to use Johnson regardless of his experience or skill level. Therefore, it seems that natural instinct and style can only get a running back so far in the modern NFL game.

Today’s athlete has everything he has ever done on a sports field measured and critiqued. No longer can an NFL team draft a running back just by watching him play or working him out. No informed decision on who to draft is made without looking at a player’s NFL combine, college and occasionally high school statistics. This paper has attempted to examine the model of the important analysis that goes on in the minds of NFL recruiters, who are searching for the next Pro-bowl running back. Because these minds are always changing, ultimately this type of statistical study is never be finished.

**Conclusion:**

In terms of per-game attendance, the NFL is the most popular professional sports league in the world. The Super Bowl is the most watched annual event on U.S. television. The NFL has the highest average team payroll of any professional sport and a salary cap that will exceed $100 million for the first time under the new collective bargaining agreement with the NFL’s player union. With all this money flowing in and out of professional football, it is important to understand the supply and demand decisions that are made for incoming NFL players. Players’
choices to train hard in hopes of making it to the NFL can ultimately reward them through multimillions dollars contracts. It is a risky decision to set your career goals to making it as a pro athlete, but the scenario section of this paper hopes to evaluate this risk through a careful statistical analysis and discover which situations warrant investing time, money and energy in your own physical skills.

This reward is even more heightened when a running back is drafted as a top 50 overall draft selection. The findings of this paper show that among the top 50 draft picks, a change in draft pick has a much greater effect on income than it does for running backs drafted after the 50th pick. Unfortunately, the forecasting power of this model is not very strong for players from many years in the past. This paper’s model over estimates the income’s of past players, probably due to the new-age marketing and fanaticism that is associated with the NFL today. This technological age allows the players to get paid in multimillion dollar contracts, with even average players getting paid more than the best players in the NFL did a couple years ago. In an attempt to avoid multicollinearity and add more variables that can help describe the income of NFL running backs, this paper created a model that is based on the average speed of the players in the three sprints at the NFL combine. This model could be interesting for future studies, but currently is not as descriptive as the final model for this paper is.

This paper hypothesized that the 40-yard dash relative to the player’s weight would have the largest effect on how running backs would get drafted in the NFL. This hypothesis proved to be correct, with the 40-yard dash being statistically significant in the final model, when used in linear form, quadratic form, and as an interaction term with weight. This shows that the attribute NFL teams are most concerned with in running backs is their ability to create momentum. Basic
physics states that momentum equals velocity multiplied by mass. The 10-yard and 20-yard dash are not long enough to let running backs hit their top velocities, and thus NFL scouts rely on the combination of 40-yard dash times (velocity) and the player’s weight (mass) when analyzing draft position.

It is important to remember that this paper is a precursor to future studies and, as such, many questions remain unanswered. However, this analysis has illuminated a few of the most important determinants of draft position for running backs in the NFL. This inquiry will hopefully lead to the further examination and a better understanding of what gives young running backs the greatest edge in the most popular sports league in America.

Foot Note: NFL collective bargaining agreement and attendance facts offered at “Reference 8”
References

<http://sports.espn.go.com/nfl/player/profile?playerId=>


Reference 3: Vance, Lloyd. “2009 NFL Combine” (February 19, 2009)  


<http://www.rotoworld.com/content/playerpages/player_main.aspx?sport=NFL&id>

