Organizing employees into teams increases productivity but also generates incentives to shirk. Recent research suggests that peer enforcement plays an important role in deterring shirking in teams. We analyze 10 years of performance and compensation data for NFL offensive linemen, a high-skill, high-salary and repeatedly interacting team, using the Hausman-Taylor estimator to control for unobservable individual-specific heterogeneity. We find evidence that teammates’ effort signals reduce the salaries of individual offensive linemen, providing an optimal, low powered sanctioning mechanism for individual workers in this setting, and that a separate, independently monitored individual effort signal also reduces salaries.

JEL Codes: J24, J3, J44, L83

Keywords: peer enforcement, teams, shirking, incentives

Introduction

Firms often organize employees into teams. Economists posit that firms organize workers into teams to take advantage of complementarities, increasing the productivity and output of teams beyond that of individual workers. Organizing workers into teams also generates problems. Holmstrom
(1982) showed that groups of workers with inter-related productive inputs generates a moral hazard problem, in the form of an incentive for some team members to supply less effort, or shirk. Successful teams exploit complementarities and reduce the incentive for individual workers to shirk.

The incentive to shirk in teams can be mitigated in a number of ways. Che and Yoo (2001) analyzed incentives in teams when the team members repeatedly interact and observe the behavior of other team mates. Their model predicts that compensation schemes that reward an employee when co-workers perform well, and punish an employee when co-workers preform poorly, have desirable properties in a workplace featuring repeated interaction among teammates. Ishida (2006) generalize this model to the case where compensation is relative, and not absolute, and shows that similar predictions emerge from the model. The key feature in these models is the existence of sanctions imposed by team members for past behavior that can be used to reduce shirking. Kandel and Lazear (1992) make a similar point, based on a model of partnerships with peer-pressure. They point out that in order for sanctions to reduce the incentive to shirk in a team, an individual team member’s “effort must affect the well-being of the rest of the team for them to have incentive” to sanction another teammate.

We develop evidence from the field that incentives like those described in the model developed by Che and Yoo (2001) exist. Mas and Moretti (2009), Ichino and Maggi (2000), and Chan et al. (2014) develop evidence of peer interaction in teams using data from cashiers in a national supermarket chain, an Italian bank, and cosmetic sales in a department store in China, respectively. Mas and Moretti (2009) focus on peer interaction and productivity; Ichino and Maggi (2000) focus on the incidence of shirking. Depken and Haglund (2011) also develop evidence of peer effects in team sports, although this evidence is based on observed productivity, not earnings of team members. We analyze the performance and compensation of members of a high-skill, high-salary team that interacts repeatedly, offensive linemen in the National Football League (NFL). Offensive line play in the NFL is complex, and the actions of individual linemen are highly inter-related. These high-skill workers interact hundreds of times over the course of a season, much like the infinitely repeated game in the model developed by Che and Yoo (2001).

We find that, after controlling for unobservable worker-specific heterogeneity, experience, and other personal and team-specific characteristics that affect performance and compensation, a specific observable effort signal generated by team mates reduces the salaries of individual offensive
linemen, but has no effect on the salaries of the team mates, providing an optimal, low powered sanctioning mechanism for individual workers. This supports the predictions from the model developed by Che and Yoo (2001). We also find that another observable signal related to shirking that is monitored by independent evaluators affects only the shirking employee, suggesting that NFL teams also use standard financial incentives to deter shirking in this setting.

Motivation and Context

Che and Yoo (2001) developed a model of team behavior that includes long-term interaction among team members; the model features an infinitely repeated game in which team members observe the effort of other workers on the team and decide how much effort to supply in each period. Ishida (2006) extended this model to the case where only relative performance analysis takes place. Each workers’ strategy is a function that maps all possible past effort decisions into a probability distribution over current effort decisions by team members. In this model a firm hires two identical workers to repeatedly perform a specific project or task. Each worker makes a binary effort decision $k$ to either supply effort (“work,” $k = 1$), or not supply effort (“shirk,” $k = 0$) in each period. Effort requires a cost $e$ but shirking entails no effort-related cost. The key feature of the model is that the workers experience close interaction in each period, so they perfectly observe each others effort decision. The model includes mutual monitoring, an important feature of any team-based work arrangement. The workers interact only through their effort decisions in this model, they cannot exchange side payments. The model predicts that effort decisions generate self-enforcing incentives that take the form of punishing other team members by shirking.

Firms do not observe individual effort decisions made by workers in this model. Instead, firms receive a binary signal, $x_i$, that is either good ($x_i = 1$) or bad ($x_i = 0$). This imperfect signal about the workers’ effort decision reflects both the individual’s effort and a random environmental shock that effects both workers. This common environmental shock is either favorable or unfavorable, and the probability of a favorable common shock is $\sigma$. In an alternative version of the model, the firm receives a single imperfect signal that is the result of the team effort and the environmental shock instead of a signal for each team member. The predictions of the model do not depend critically on the nature of this signal.
The team production arrangement is open ended and terminates with probability $1 - \delta$ at the end of each period, so $\delta$ indicates how long teams remain together. The workers always have incentive to shirk since effort is costly but shirking is not easily observed given the imperfect effort signal. The firm wants to induce workers to supply positive effort in each period, because the outcome when both workers provide positive effort is more valuable than other outcomes. The firm’s problem is to motivate both workers to provide positive effort in every period at minimum cost using some wage scheme. Specifically, the wages cannot be negative and depend on the verifiable signals received by the firm. Clearly, an important feature of the wage scheme designed by the firm in this context is how an individual’s compensation relates to the performance of other team members.

Two possible wage schemes exist in this setting: a wage scheme based on relative performance evaluation (RPE) under which a worker is penalized when other team members perform well and rewarded when other team members perform poorly; and a wage scheme based on joint performance evaluation (JPE) under which a worker is rewarded when other team members perform well and penalized when other team members perform poorly. Tournament theory represents a special form of RPE; under RPE, the worker perceived to perform the best, based on the signals received, earns the highest compensation, the second best the second highest, and so on. RPE wage schemes appear to be ineffective when workers interact closely with each other, perhaps because it leads to competition among team members, especially when workers can sabotage others in the competition [Lazear (1989)]. JPE wage schemes overcome these problems in settings where workers interact repeatedly by reducing the negative aspects of competition while providing team members with a mechanism to deter shirking. Under JPE, a worker can shirk in order to punish another team member because the other team member’s compensation will be reduced by this action. Che and Yoo (2001) show that the firm can make a larger profit under JPE than under RPE when workers repeatedly interact, because the total cost of the wage scheme is lower under JPE than RPE.

Che and Yoo (2001) demonstrate that JPE wage schemes are optimal relative to RPE wage schemes in settings where workers on a team interact repeatedly, no matter how long workers remain in teams. RPE wage schemes are more likely to be optimal as the probability of favorable common shocks increases. JPE wages schemes have another important property in this model: the explicit incentives, those that affect wages directly, must be relatively low powered for shirking
to be a credible punishment strategy. In other words, for shirking to be deterred, the deterrence mechanism must affect the other worker more than the worker who seeks to deter shirking; the effect must be asymmetric.

The presence of JPE or RPE wage schemes in a team can be empirically investigated. Given data on the earnings of individual workers who are organized into teams and interact repeatedly in production, and proxy variables for positive or negative effort signals received by firms, under RPE wage schemes, a larger number of negative signals from other workers on the team would be associated with higher earnings by a given worker, and a larger number of positive signals from other workers would be associated with lower earnings for a given worker. Under JPE wage schemes, a larger number of negative effort signals from other workers represent peer sanctions. In addition, the model developed by Che and Yoo (2001) predicts that JPE wage schemes would be more likely to be observed in a setting where the teams are relatively long-lived and where the probability of common positive shocks to team members is relatively low.

One setting where JPE wage schemes might be present is among offensive linemen in professional football leagues like the NFL. Note that we interpret the offensive line as a group of workers (football players) organized into a “team” within a professional football team. Che and Yoo (2001) identify three characteristics of a team in the context of their model: (1) frequent and consistent interaction over a long period of time, (2) autonomy to make independent decisions on assignments and problem solving; and (3) encouragement to monitor and motivate each other. Offensive linemen must work together closely in a highly coordinated way in order to create running room for ball carriers when the team runs the ball and to protect the quarterback when the team passes the ball. The assignments for individual linemen are highly inter-related and complex. NFL teams attempt to reduce turnover in personnel on the offensive line, and since NFL players only gain free agency after three years of experience in the league, offensive linemen often play together for a number of seasons. During the game, the offensive line makes its own play calls at the line of scrimmage after the quarterback calls the offensive play that will be run in the huddle, suggesting that the offensive line has some independent decision making power. And anyone who has watched an NFL game would agree that offensive linemen continually monitor and motivate each other.

On an NFL team, the offensive line typically contains five players, or in this context, team members. The center plays in the middle of the line and snaps the football, handing or throwing it
between his legs to the quarterback to initiate each play. Next to the center are the right and left guard, and outside the guards are the right and left tackle. Offensive linemen cannot touch the ball in most circumstances; the goal of this team is to provide open space for other players advancing the ball by running and to protect the quarterback when he passes the ball. The primary activity performed by offensive linemen is called “blocking.” This activity involves physical interaction with defensive players and a large body of rules specify exactly how an offensive lineman can legally touch a defensive player and also specify illegal forms of interaction that, if detected by an official, will result in a penalty in the form of moving the ball backward by a specific distance, usually five or ten yards. Offensive linemen also must, by rule, remain absolutely still until the center snaps the ball to the quarterback and cannot run down field on passing plays.

In the context of effort supply, we interpret penalties committed by offensive linemen as a signal of shirking. Seven officials monitor play in the NFL. Three of these officials monitor offensive linemen for rule violations, as these workers are subject to a large number of rules that proscribe many specific activities. Offensive linemen can commit a number of infractions that result in a penalty if observed by an official. Two common penalties committed by offensive linemen are false starts and holding. A false start occurs when an offensive lineman moves before the center has snapped the ball to the quarterback and results in a five yard penalty. Holding occurs when an offensive lineman (or other offensive player) grabs or tackles a defensive player in a way prohibited by the rules and results in a ten yard penalty. When an offensive lineman commits holding, he supplies less effort than he would do if no penalty was committed. A false start is a mental error, but can also be interpreted as supplying less effort, in this case effort to remember the signal for the snap of the ball.

In the model developed by Che and Yoo (2001), firms receive a signal that reflects both effort supplied by workers and a common random shock. NFL teams receive signals about the effort supplied by offensive linemen. Clearly, NFL players could engage in a large number of types of shirking. In addition to the penalties discussed above, we also consider a second type of signal about effort supplied by offensive linemen, giving up sacks. On a passing play, the goal of the offensive line is to keep the defensive players from tackling the quarterback before he can throw the ball down field. A sack occurs when a defensive player tackles the quarterback in the backfield before he can throw the ball. This results in a loss of yardage, a bad outcome for the offense. We interpret
a sack allowed by an offensive lineman as a signal of shirking. Implicitly, if an offensive lineman would have supplied more effort, a sack could have been avoided on a play. Unlike penalties, which are monitored by the officials, coaches and other workers must monitor offensive linemen on passing plays to determine the amount of effort supplied.

Since the effort signal includes a random shock component, shirking can be somewhat masked by ability, random events, or the presence of a strong opponent. Shirking usually occurs when player is facing an evenly-matched or stronger opponent (a negative shock), since shirking would be easily observed if facing weaker opponent (a positive shock). In practice, shirking may be determined by a coach after reviewing the game video, which is costly and inefficient. It also cannot be easily determined during a game, when coaches must make personnel decisions based on the game situation and specific player match-ups. Also some sacks can’t be awarded to specific players even after video review. But offensive linemen on the field during the game have better information about effort supply. Because of repeated interactions, they know more than anyone else about their colleagues’ effort supply. So mutual monitoring and punishment would be an efficient and less costly way to deter shirking during games.

We empirically analyze the relationship between these two effort signals and the salaries earned by these players. Again, we assume that penalties committed and sacks allowed can be interpreted as signals of effort supplied by offensive linemen. Since we observe both signals received by the firm, and the salaries of the team members, this information can be exploited to determine if a JPE wage scheme exists for these high-skilled workers. Again, under JPE wage schemes, workers are penalized for poor performance by team mates, a mechanism through which workers, over the course of repeated interaction, punish other workers for supplying lower levels of effort. If penalties committed, or sacks allowed by team mates reduces the compensation of offensive linemen, then outcomes in this setting are consistent with JPE wage schemes, supporting the model developed by Che and Yoo (2001).

Data Description

Our data include information about the performance and earnings of NFL offensive linemen, and their teams, over the 2000 to 2009 regular seasons. Our basic unit of observation is an individual
NFL offensive lineman over a season. Offensive line play is highly inter-related. Unlike many other positions on an NFL team, the offensive linemen must cooperate and work as a unit in order to perform well. The data set we construct contains five types of information about offensive linemen: (1) individual performance data for each season, (2) player characteristics; (3) salary, (4) financial data for NFL teams and (5) team offensive performance. The data were collected from a number of sources. The player performance and characteristics data are from several sports data websites, including the official NFL website, Stats Incorporated’s fee-based STATSPASS database, and www.pro-football-reference.com; the salary data are from the USA Today NFL Salary database; the team-specific financial data are from the estimates published in Forbes magazine and on their website, and the team’s offensive performance data are generated from data on www.pro-football-reference.com. There were 2652 player-year observations including data from 688 unique players over the 2000 to 2009 seasons. Summary statistics for the player characteristics and salary variables are shown on Table 1.

NFL player compensation consists of guaranteed pay in the form of a “signing bonus” an annual “base salary” paid to players who remain on the team’s roster, and performance-based pay that depends on team performance and the number of plays each player participated in over the course of a season. Neckermann et al. analyze such bonus pay in a different setting. Our measure of salary is the “cap value” accounted for by each player on the offensive line. The NFL regulates the fraction of total revenues that can be paid to players. While this is often called a “salary cap”, a more accurate description is a ceiling and floor on total payroll. The NFL “salary cap” does not regulate or limit the amount that can be paid to any player; it places a lower and upper limit on the total payroll for players on each team as a fraction of certain revenues earned league wide. Most salaries in the NFL are not guaranteed, and players under contract for multiple years can be released at any time. The “cap value” is the compensation paid to each player that counts toward total team payroll; it includes base salary, a prorated portion of the signing bonus, and incentive bonuses. The “cap value” thus reflects the total compensation of the player in a season, including performance bonuses that depend on the number of plays participated in per season. Cap value can vary depending on the number of plays an individual participates in, so coaches can affect this

---

1Beginning in the 2002 season, NFL players qualified for performance-based compensation based on playing time. All players qualify for this performance based pay. The largest bonus was $42,048 in 2002 and increased to $299,465 in 2012.
salary by keeping players on the sideline. Cap value is the standard salary measure in the NFL and has recently been used by Berri and Simmons (2009), Simmons and Berri (2009), Berri et al. (2011) and Keefer (2013). The average salary in the sample is $1.87 million per season. The salary variable exhibits quite a bit of variability and has a long right tail.

In football, there are three offensive line positions: center, guard, and tackle. In our data, some players are identified as generic offensive linemen and not as playing a specific position. Players identified as generic offensive linemen either played multiple positions on the offensive line, or information about their specific position is not available. The latter case happens in the early years of our data. We create 4 indicator variables to identify linemen playing these 4 positions. Each indicator equals 1 when the player plays that position. From Table 1, about one third (35%) of the players are identified as generic offensive linemen in the sample. Notice that creating dummy variables in such way has a limitation. We don’t differentiate between left tackle and right tackle here. NFL teams often treat left tackles and right tackles differently. A left tackle is often paid more because he protects the quarterback’s “blind side”; since most NFL quarterbacks are right handed, a defensive player approaching the quarterback from his left cannot be easily seen.

Size is an important characteristic for offensive linemen. We collected data on the players height in inches and weight, which ranges from 228 to 375 pounds. The offensive linemen in the sample have a large weight range; the difference between the maximum and minimum weight is 147 pounds. However, weight may not be a good indicator of characteristics which affect performance in this setting.

Like other North American professional sports leagues, the NFL conducts an annual reverse-order entry draft to allocate incoming players to teams. In this sample period, the NFL entry draft consisted of seven rounds. Overall Draft Selection is the position where each player was selected in the draft. Players are selected in approximately their order of expected ability. Hendricks et al. (2003) show that draft order is an important measure of the value of a player in the NFL. If a player was not drafted, teams expected that he might not be good enough to make an NFL roster. For undrafted players, we assign 277 to their pick number\(^2\). We also incorporate into our model an indicator variable Undrafted, equal to one of the player was not drafted, together with Overall Draft Selection, to capture the effect of not being drafted on salaries, because the difference between the

\(^2\)The largest draft number for drafted players in the sample is 276.
last picked player and undrafted player may not be simply measured by one position in the draft order. 20% of the players in the sample were undrafted.

Better players should start in more games in a season, and receive higher salaries. The variable *Games Started* equals the number of games started by each player in the sample. The average offensive lineman in the sample started ten games per season. *Years Experience* measures the number of years played in the NFL. For players who entered the NFL through the entry draft, experience is based on the draft year. For players who were not drafted, experience is based on the first year they entered the league.

The variables of interest are the signals of player effort. Stats Incorporated’s STATSPASS database contains information about the performance of individual offensive linemen in the NFL. This includes information about the number of penalties committed by each offensive lineman, the number of sacks given up by each offensive linemen, and the yards lost associated with each of these effort signals.

For each player, we divide *penalty yards* by *penalties committed* to estimate *Yards per Penalty*. Average yards per penalty committed reflects both the rate at which penalties are committed and how much the team is punished for the penalties. For example, if two players have the same total penalty yards, say 15, three 5 yard penalties (average 5) may be less harmful to the team than one 15 yard penalty (average 15). The second reason to use average yards per penalty is that reserves play fewer downs than starters, and could have fewer penalties and fewer penalty yards. But a reserve player’s average yards per penalty may not be small. The same problem also applies to sacks allowed and sack yards. Thus we also calculate *Yards per Sack* from *sacks allowed* and *sack yards*. Moreover, since offensive line play is highly inter-related, we incorporate other offensive linemen’s performance variables into the model. Those variables are *Teammates’ Yards per Penalty*, and *Teammates’ Yards per Sack* on Table 1.

Manski (1993) pointed out a problem that occurs when a researcher observes the distribution of behavior in a population and tries to determine the extent to which this behavior influences the behavior of individuals in this population, the “reflection” problem. We avoid this issue by explaining variation in the salary of an individual offensive lineman using variables that reflect the behavior of all other members of the offensive line excluding that individual. *Teammates’ Yards per Penalty* and *Teammates’ Yards per Sack* reflect the behavior of other teammates and differs
The average offensive lineman in the sample committed 3.29 penalties per season. In only about 17% of the player-seasons a lineman did not commit a penalty. However, many of those penalty-free player-seasons were by little-used reserves. Among players who started 8 or more games, only about 4% of the player-seasons were penalty free. The maximum number of penalties committed in the sample was 17. The average yards per penalty in the sample was just over 6.

The average offensive lineman in the sample allowed just under three sacks over the course of a season. In about 18% of the player-seasons the player did not give up a sack. Again, this is sensitive to the number of games started. Among players who started at least 8 games in a season, the average number of sacks allowed was four, and only 3% of the player-seasons in the sample were seasons with no sacks given up. The average yards per sack given up in the sample was just over 5.

Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Salary (millions)</td>
<td>1.87</td>
<td>1.89</td>
<td>0.089</td>
<td>14.20</td>
</tr>
<tr>
<td>Position: Center</td>
<td>0.17</td>
<td>0.38</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Position: Guard</td>
<td>0.21</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Position: Tackle</td>
<td>0.27</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Multiple Position Player</td>
<td>0.35</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Height (inches)</td>
<td>76.4</td>
<td>1.63</td>
<td>72</td>
<td>81</td>
</tr>
<tr>
<td>Weight (pounds)</td>
<td>310</td>
<td>20.2</td>
<td>228</td>
<td>375</td>
</tr>
<tr>
<td>Years Experience</td>
<td>5.75</td>
<td>3.32</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Overall Draft Selection</td>
<td>137</td>
<td>97</td>
<td>1</td>
<td>277</td>
</tr>
<tr>
<td>Undrafted player</td>
<td>0.21</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Games Started</td>
<td>10.1</td>
<td>6.45</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Yards per Penalty</td>
<td>6.02</td>
<td>3.31</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Yards per Sack</td>
<td>5.28</td>
<td>3.22</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Teammates’ Yards per Penalty</td>
<td>7.18</td>
<td>0.77</td>
<td>4.93</td>
<td>10</td>
</tr>
<tr>
<td>Teammates’ Yards per Sack</td>
<td>6.41</td>
<td>0.94</td>
<td>3</td>
<td>10.37</td>
</tr>
</tbody>
</table>

**Empirical Approach**

We analyze longitudinal data on the performance of NFL offensive linemen over ten seasons. We observe individual performance and salary for the same offensive linemen over multiple seasons, so our data constitute an unbalanced panel. Controlling for unobservable heterogeneity, in the form
of ability, desire, and other intangible characteristics, represents an important econometric issue in this setting. Researchers typically consider two types of estimators when analyzing panel data with unobservable heterogeneity: fixed effects (FE) and random effects (RE) estimators. Mundlak (1978) points out that the RE estimator assumes all explanatory variables are exogenous and uncorrelated with unobservable individual effects, while the FE estimator assumes all explanatory variables are endogenous and correlated with the unobservable individual effects. Therefore, in the case where some explanatory variables are exogenous and uncorrelated with unobservable individual effects and other explanatory variables are endogenous and correlated with unobservable individual effects, neither the FE nor the RE estimator works well. In this setting, we likely face a situation where some, but not all of the observable variables are correlated with unobservable individual characteristics. Kahn (1993) discusses this problem and the merits of these two estimators in the context of empirical research on earnings in professional sport.

Hausman and Taylor (1981) proposed an estimator appropriate for the case where some of explanatory variables are correlated with unobservable individual effects and others are not. This estimator is based on an instrumental variables (IV) approach. The instruments are constructed using the strictly exogenous variables’ between and within variations. This estimator is called the Hausman-Taylor or HT estimator [Baltagi et al. (2003)]. Buraimo et al. (2008) applied the HT estimator to data from professional sport; Dixit and Pal (2010) used it to analyze group incentives.

In general, the estimator developed by Hausman and Taylor (1981) takes the form

\[ y_{it} = X_{it}\beta + Z_i\alpha + d_i + u_{it} \]  

(1)

where the subscript \( i \) identifies the cross-sectional unit \( (i = 1, 2, \ldots, N) \), and the subscript \( t \) identifies the time periods \( (t = 1, 2, \ldots, T) \). In this context, the cross-sectional units are players and the time periods seasons. \( X_{it} \) is a vector of time-varying explanatory variables and \( Z_i \) is a vector of time invariant explanatory variables. \( d_i \) and \( u_{it} \) are unobservable random variables that affect the dependent variable \( y_{it} \). \( d_i \) is assumed to be \( i.i.d.(0, \sigma^2_d) \) and \( u_{it} \) is \( i.i.d.(0, \sigma^2_u) \). Both are independent of each other and among themselves. \( d_i \) is the unobservable individual effect; in this setting, \( d_i \) captures a players skill, ability, motivation, or “coachability” that can not be reflected in performance statistics and other unobservable factors that affect the performance of offensive linemen in
the NFL, including but not limited to the will to win, morale, leadership, and other intangibles. Both $X$ and $Z$ can be split into two sets of variables, i.e. $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$. $X_1$ and $Z_1$ are exogenous and uncorrelated with both $d_i$ and $u_{it}$; $X_2$ and $Z_2$ are endogenous and correlated with $d_i$ only.

Again, neither the RE nor the FE estimator applies in this setting. The RE estimator, which is basically GLS applied to Equation (1), ignoring the endogeneity of $X_2$ and $Z_2$, will yield consistent but biased parameter estimates, while the FE estimator, or mean-differencing the explanatory variables in Equation (1), eliminates the individual effect $d_i$ as well as the time invariant variables $Z_i$ and hence cannot yield estimates of the vector $\alpha$, though it can yield consistent estimates of $\beta$. The HT estimator resolves this problem.

Hausman and Taylor (1981) develop a variant of the standard IV estimator. It first premultiplies Equation 1 by $\Omega^{-1/2}$, where $\Omega$ is the covariance matrix of error term $d_i + u_{it}$. After the transformation, it uses a standard two stage least squares IV (2sls) approach with instruments $[Q, X_1, Z_1]$, where $Q$ is for demeaning the variable, specifically $Q y_{it} = y_{it} - \bar{y}_i$. Therefore, the Hausman-Taylor estimator is basically equivalent to perform 2sls using $[\tilde{X}, \tilde{X}_1, Z_1]$ as instruments. Intuitively, $\tilde{X}$ can be instrument for $X_2$ and $\tilde{X}_1$ is for $Z_2$. The advantage of the HT estimator is that all the instruments are derived from within the model. $\tilde{X}$ is the matrix of the deviation of $X$ (both $X_1$ and $X_2$) from its associated mean. $\tilde{X}_1$ is the mean of the exogenous time-varying variables. These are the standard within and between components of the FE and RE estimators.

Another important issue associated with the HT estimator is the identification condition for the model. As pointed out by Baltagi et al. (2003), if the number of exogenous time-varying variables $X_1$ is greater or equal to the endogenous time-invariant variables $Z_2$, the model is identified and the HT estimator is more efficient than the FE estimator. If the condition fails, the model is under identified and the HT estimator cannot outperform FE and the coefficient on $Z$, $\alpha$, cannot be estimated.

The specific form of the HT estimator in this case is:

$$\ln salary_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + Z_{1i}\alpha_1 + Z_{2i}\alpha_2 + d_i + u_{it}$$  (2)

where $X_{1it}$ is a vector of $n \times k_1$ exogenous, time-varying explanatory variables that are assumed
to be uncorrelated with $d_i$; $X_{2it}$ is a vector of $n \times k_2$ endogenous, time-varying variables that are assumed to be correlated with $d_i$; $Z_{1i}$ is a vector of $n \times l_1$ exogenous, time-invariant explanatory variables that are not correlated with $d_i$; $Z_{2i}$ is a vector of $n \times l_2$ endogenous, time-invariant variables that are correlated with $d_i$; and $d_i$ is a player-specific effect. The model is identified when $k_1 \geq l_2$.

Note that $X_{1it}$ contains a vector of variables identifying the team that each lineman played for over the sample period. This variable captures unobservable heterogeneity across teams. It is time varying because some players change teams in the sample period. Unobservable team specific heterogeneity may arise from various sources, such as culture, managerial style, owner and coach preferences and factors related to the city where team plays. The general business environment, stadium deal, and residents’ enthusiasm for football in the city could all be captured by this team effect.

Again, we have unbalanced panel data; the cross-sectional unit is an individual player. The dependent variable is the natural logarithm of the real salary in each season. The model contains 43 variables classified in four categories (exogenous time-varying, endogenous time-varying, exogenous time-invariant and endogenous time-invariant). The vector of exogenous time-varying variables contains 34 variables, including 32 team dummy variables $Years~Experience$ and its squared term, $Years~Experience^2$, which together measure the effect of experience. All these variables are assumed to be independent of the unobservable individual effects. It is unlikely that unobservable individual effects (for example, ability, will to win or drive) influence the player’s team or experience. Though a talented player is expected to play longer in the league, a lot of geniuses regrettfully retired early because of injury. The player career is short, normally less than 10 years. Some players without great talent can also stay in the league for a relative long time as long as they can keep healthy. We add both $Years~Experience$ and $Years~Experience^2$ into the model following the same logic of the quadratic relationship between salary and age, and hence we expect a positive coefficient on $Years~Experience$ and negative coefficient on the squared term.

The vector of endogenous time-varying variables includes 6 variables: an indicator variable identifying linemen who played multiple positions over the course of the season, the number of games started in the season, and four variables reflecting the effort signals received by the team: yards per penalty, yards per sack allowed, teammates yards per penalty, and teammates yards per sack allowed. Linemen who played multiple positions are likely to be reserves filling in occasionally
at many positions, perhaps because they lack the ability to start at a position full time. We expect the coefficient on this variable to be negative. Better players should start in more games in a season, so *Games Started* should have a positive coefficient.

Again, signals received about individual players (yards per penalty and yards per sack) reflect both effort decisions and a common shock experienced by all offensive linemen on a team. Penalty yards and sack yards reflect shirking by a player, so those variables are expected to have negative coefficients, based on standard efficiency wage theory. The signals received about other players effort (teammates’ yards per penalty and teammates yards per sack) are motivated by the model of repeated interaction among workers on a team developed by Che and Yoo (2001) and discussed above. The estimated parameters on these variables will indicate the presence of JPE wage schemes, if negative, and RPE wage schemes, if positive.

The performance variables are all likely to be correlated with the unobservable individual effect. For example, if a player works very hard in both practice and games, he will probably perform well in games and the good performance would in turn affect his desire to win and/or other unobservable factors.

The only exogenous time-invariant variable is *Height*, which is constant over time and can not be affected by unobservable individual effects. Conversely, it is also unlikely that height would influence ability or other unobservable individual effects. Height may be a basic physical qualification to become an offensive lineman. But conditional on becoming a linemen, we can’t say taller players are superior to shorter players and the players themselves would not believe that being a few inches taller would bring them some advantage over other players. The unobservable individual heterogeneity should be uncorrelated with height. However, among offensive linemen, generally speaking, tackles are tallest, then guards, and centers are often the shortest players. Also the teams usually pay tackles, especially left tackles, more. So it seems taller players earn more money and hence we tentatively expect a positive coefficient on *Height*. Note that in our data, about one quarter of players switch positions over their careers. So height and position are not highly correlated.

The two endogenous time-invariant variables are *Undrafted* and *Overall Draft Selection*. These two variables, especially the second one, are regarded as important indicators of players’ ability and hence are expected to have significant effects on salary. Moreover, since draft order reflects
individual ability, this variable is likely to be correlated with the unobservable individual effects. A player who is believed to have greater ability and potential will be selected earlier in the draft. Those factors will in turn influence attitude toward the game and effort. Since a smaller draft number means a better player, we expect Overall Draft Selection will have a negative coefficient. But for Undrafted, the sign could be either negative or positive, because the effect of being undrafted is captured by both these variables. If the sign is positive for Undrafted, the coefficient on Overall Draft Selection plus the coefficient on Undrafted should be negative, in which case undrafted players earn less than the last drafted player.

In our model, $k_1 = 34$ and $l_2 = 2$, even after omitting one team indicator variable, $k_1$ is still much larger than $l_2$. Therefore, our model is identified.

**Results and Discussion**

The parameter estimates, standard errors and P-values obtained from the HT estimator applied to Equation (2) are shown on Table 2. Note that we do not report the parameter estimates from the 31 team indicator variables since their effects are not our primary interest here. These results are available by request from the authors.

The relationship between experience and salary takes the standard hump-shaped form, first increasing as human capital accumulates and then decreasing as it depreciates. Declining physical ability, and the cumulative effect of injury and physical wear-and-tear also contribute to the decline in earnings as experience increases in this setting. As expected, linemen who play multiple positions earn less than those who play a single position, even holding games started constant. This effect may also reflect returns to specialization among offensive linemen in the NFL. Starting linemen earn a significant premium over reserves who start no games, or perhaps only a few games, per season.

The variables of interest are the effort signals for the two effort signals, committing penalties and allowing sacks. Again, the salary variable reflects a signing bonus, base salary, and performance related pay that varies systematically with the number of plays participated in. Effort signals can affect salary if coaches make substitutions based on these signals over the course of a game, generating differences in the number of plays each player participates in over the course of a season.
The results exhibit an asymmetric pattern that supports the idea that a JPE wage scheme exists in this setting, as one effort signal from teammates, giving up sacks, affects the salary of other players. Each additional yard per sack given up by teammates reduces the earnings of an individual NFL offensive lineman by about 2.5%. However, yards per sack given up by an offensive lineman do not reduce his own earnings. This pattern of results suggests that a low powered mechanism exists through which offensive linemen can punish teammates for shirking. In the repeated interaction that takes place on an NFL offensive line over a season, if teammates observe another lineman shirking, they can punish him by giving up sacks without reducing their own salary, providing a low power mechanism for enforcing the a sub-game perfect equilibrium in which these workers supply effort in each period, just as the model developed by Che and Yoo (2001) predicts. Note that sacks have less impact than offensive holding, a penalty that might be committed to avoid a sack; offensive holding results in a 10 yard penalty while the average sack in the 2002-2010 NFL seasons resulted in a 6.37 yard loss (median 7 yard loss). This result is consistent with the use of JPE wage schemes based on sacks allowed among teams of workers on the offensive line.
The other effort signal, penalties committed by players, does not appear to be a viable mechanism to punish other workers who shirk. The yards per penalty committed by other teammates does not affect the earnings of individual linemen in this sample. However, the larger the yards per penalty committed by an NFL offensive lineman, the lower that lineman’s salary, other things equal. Recall that, unlike sacks allowed, the commission of penalties in the NFL is monitored by an independent party, the team of referees that officiate NFL games. Each additional yard of penalties committed reduces the earnings of the offensive lineman by 1%. This result can be motivated by the model developed by Kvaløy and Olsen (2012), which includes indispensable human capital in teams. Alternatively, standard efficiency wage models, or principal-agent models with monitoring by the firm, in which firms penalize workers who shirk with lower wages if they can observe this shirking, can also motivate this result.

The results indicate that a salary premium to height exists in this setting. Taller offensive linemen earn more than their shorter teammates, other factors constant. The premium to height is substantial, about 15% per inch. The premium could be due to higher productivity among taller linemen. Persico et al. (2004) find a wage premium to height in the general population and attribute this to events in early adulthood. Schultz (2002) finds a wage premium to height and attributes it to human capital accumulation. These two explanations could also hold in this population.

The results also show that linemen drafted higher in the entry draft earn a higher salary, even years after being drafted and controlling for experience. This persistence of draft position in salaries was also documented by Hendricks et al. (2003) in the NFL. Note that 65% of the variation in the dependent variable can be attributed to unobservable player-specific heterogeneity captured by the random effects term in Equation 1.

Robustness Checks

Table 2 contains results for a single model specification. The HT estimator can be sensitive to the specification of the time-varying endogenous and exogenous variables. No reliable, commonly used tests exist to provide guidance about which variables belong in which category in the HT estimator. We performed a number of robustness checks on the results reported on Table 2. The results, specifically the sign and significance of the effort signal variables, were not sensitive to these changes to the model.
We added a series of indicator variables for each season in the sample, to capture any unobservable heterogeneity in offensive line play that would affect compensation, sacks allowed and penalties committed. These factors could include changes to the specific rules about what offensive linemen can and cannot do, and any changes in the enforcement of the existing rules that might vary systematically across seasons. The inclusion of this vector of indicators to the time varying exogenous variable list had no effect on the parameter estimates of interest. We also added an indicator variable for changes in the head coach of the team; this also had no effect on the results.

Equation (1) does not control for systematic variation in the salaries of offensive linemen across positions on the line. If sorting by ability takes place, say the most talented linemen become tackles, or the smartest linemen become centers, then this needs to be controlled for in the regression model. We included a vector of variables identifying the specific position played by the linemen in our sample, both as an time varying exogenous and time varying endogenous variable. The inclusion of these indicator variables had no effect on the results. Similarly, changing the indicator variable for a lineman who played multiple positions from time varying endogenous to time varying exogenous had no effect on the results.

We also added BMI to the model as a time-varying endogenous variable. Although height does not change, weight does. BMI could affect agility, or other factors that affect the play of a lineman. Adding BMI, and BMI² had no effect on the results.

The relationship between sacks allowed and the salary of offensive linemen could be affected by some confounding factors like the tendency of the team’s offense to pass, the overall efficiency of the team’s offense, or some other characteristic of the offense that affects the team’s overall success. The results reported on Table 2 were robust to the inclusion of a variety of variables like the fraction of plays that were passes, yards per pass, the fraction of passes completed, and first down efficiency to the list of time varying endogenous variables.

Differences in team financial conditions might systematically affect the effect of the effort signal on player salaries. For example, teams with lower revenues might punish players with many negative shirking signals more than teams with higher revenues. The results reported on Table 2 were robust to the inclusion of a variety of variables reflecting the revenues earned by teams, and the fraction of the team’s revenues paid to players.

We also estimated a fixed effect model version of Equation (1); the results were similar to the
Hausman-Taylor results on Table 2 for the effort signal variables. Note that the 3 time-invariant variables *Height*, *OverallDraftSelection* and *Undrafted* were omitted due to collinearity. As mentioned in empirical approach section, the FE estimator can yield consistent estimates of time varying variables by eliminating the individual effect and time-invariant variables.

**Evidence From Play-by-play Data**

The regression results in the previous section suggest that giving up a quarterback sack represents a low powered effort signal in a joint performance evaluation wage scheme. In order for giving up a quarterback sack, a form of shirking, to be a low powered signal, it must be possible for NFL offensive linemen to send such signals at points in the game where the game outcome will not be too adversely affected. For example, sending such a signal when the player’s team is about to score a tying or go ahead touchdown would not be low powered, since it could lead to the team losing a winnable game. Also, not all sacks allowed are shirking-related signals; in some cases, sacks are unavoidable consequences of the game situation and personnel.

In order to assess the viability of allowing sacks as a shirking-related signal in a JPE wage scheme, we analyzed play level outcome data from all regular season games in the 2002 through 2010 seasons. The sample contains information about 305,483 individual plays. We removed kickoffs, field goal attempts, punts, extra point attempts and two-point conversion attempts, as these plays could not result in a quarterback sack. We identified plays resulting in a quarterback sack, and the number of yards lost, from the remaining plays. A sack was defined as a play in which the quarterback was tackled at or behind the line of scrimmage. Plays where the quarterback advanced the ball for positive yardage were not identified as sacks. 10,394 of these 305,484 plays resulted in a quarterback sack for an average frequency of 3.4 sacks per 100 plays. The average sack resulted in a loss of 6.37 yards; the longest sack in the sample resulted in a loss of 38 yards. Unfortunately, we cannot link specific offensive linemen to a sack in this data set. Only the season-level data set analyzed in the previous section links specific offensive linemen to quarterback sacks allowed. This data source does contain information on the exact game conditions in terms of time, field position, down and distance, and score, for each play during the NFL season.

---


4 Philadelphia quarterback Donovan McNabb versus the Oakland Raiders on 18 October 2009 in the second quarter when trailing by 7 points.
We identify games and game situations where giving up a quarterback sack would not have a large impact on the game outcome, or season outcome for teams. In these situations, giving up a quarterback sack would have a low cost to the team in terms of the effect of this signal on the game or season outcome, a low powered signal.

We identified five different situations in which a quarterback sack might not affect game or season outcomes: games early in the season, games with large point spreads where one team is much stronger than the other team, plays during games where on team has a large lead, plays late in games where on team has a large lead, and games where one of the teams has been eliminated from playoff contention. Sacks during these games, or plays during games with these conditions, could be considered low powered signals.

<table>
<thead>
<tr>
<th>Table 3: Sack Statistics for Low and High Power Game Situations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sack frequency</td>
</tr>
<tr>
<td>Yards lost per sack</td>
</tr>
<tr>
<td>Score difference ≤13</td>
</tr>
<tr>
<td>Sack frequency</td>
</tr>
<tr>
<td>Yards lost per sack</td>
</tr>
<tr>
<td>Score difference ≤13, 4th Q</td>
</tr>
<tr>
<td>Sack frequency</td>
</tr>
<tr>
<td>Yards lost per sack</td>
</tr>
<tr>
<td>Point spread ≤9</td>
</tr>
<tr>
<td>Sack frequency</td>
</tr>
<tr>
<td>Yards lost per sack</td>
</tr>
<tr>
<td>Teams with ≤9 losses</td>
</tr>
<tr>
<td>Sack frequency</td>
</tr>
<tr>
<td>Yards lost per sack</td>
</tr>
</tbody>
</table>

Table 3 summarizes the frequency and losses for quarterback sacks in each of these low and high power game situations. The first two columns are for high power game situations and the second two columns are for low power game situations. The first case is games early and later in the season. We divided games into the first 4 games of the season, and the last 12 games of the
season. Games early in the season may be less important than games later in the season. From the top panel of Table 3, 228,480 plays took place in games 5 through 16 in this sample, and 77,003 plays took place in the first four games. The frequency of sacks allowed was identical in these two periods, 3.4 sacks per 100 plays. The average loss on each sack was similar. There were 7,756 sacks in games 5-16 and 2,638 sacks in games 1-4 in these seasons.

The second low power setting is plays when one team has a large lead on the other team. We define a large lead as a lead of 14 points or more. From the second panel on Table 3, of the 305,483 plays analyzed, 60,875, just under 20% of the plays, took place when one team had a 14 point lead or larger. The sack frequency was slightly higher when one team had a large lead, and the average sack resulted in a slightly larger loss, 6.71 yards, compared to the average loss during closer game situations. A substantial number of sacks took place during these low powered game situations, providing offensive linemen with ample opportunities to send low powered effort signals. The third panel summarized sacks in the fourth quarter when one team has a large lead. The should be an even more low powered situation, since little time remains in the game for the losing team to stage a comeback. The 27,749 plays run in this condition represents 9% of the plays in the sample. Again, the sack frequency his higher and the average sack results in a larger loss in this relatively low powered situation.

The fourth low powered setting analyzed are games where one team is much stronger than the other team. Since the weaker team is less likely to win a game, games with a large point spread could represent a low powered setting since the stronger team is much more likely to win the game no matter how many sacks are given up. We obtained data on the closing point spread for all NFL regular season games over the 2002-2010 seasons. We defined games where one team is perceived as much stronger than the other as games when one team was favored by 10 or more points. This occurred in about 15% of the games played in this period. The rate of sack frequency was the same in games with a heavy favorite and games with no favored team or games with a smaller point spread. The average loss per sack was similar in the two groups of games.

Finally, games involving a team with no chance of making the playoffs could be low powered settings, since one of the teams has a reduced incentive to win the game. During this period, no team with more than 9 losses qualified for the NFL post season, so we split the sample into game involving at least one team with 10 ore more losses and games with no teams with 10 or more
losses. Games with one or more teams with 10 or more losses take place late in the 16 game NFL season. From the last Panel on Table 3, the sack frequency and average loss per sack were very similar in these two groups of games. 1,262 of the 10,394 sacks in the sample took place in games where one or more teams had 10 or more losses.

From Table 3, NFL offensive linemen have substantial opportunities to send low powered effort signals by allowing a quarterback sack, over the course of an NFL season. Thousands of quarterback sacks take place in games, or game settings, where a quarterback sack is unlikely to have a large impact on the game or season outcome for the team.

Conclusions

Che and Yoo (2001) and Ishida (2006) developed models that describe the optimal incentives in a setting where a team of workers repeatedly interact in producing a specific type of output. These models conclude that when workers on a team interact repeatedly, firms have an incentive to use JPE wage schemes, as they result in lower incentive costs when workers can monitor the effort decisions made by other team members and use a low powered punishment mechanism to deter other team members from shirking.

Offensive line play in the NFL contains a number of institutional characteristics that closely resemble the setting for the models developed by Che and Yoo (2001) and Ishida (2006). These workers interact repeatedly, can monitor each other, often motivate each other, and have significant autonomy to carry out their tasks during the course of a game; in addition, relative performance evaluation occurs in this setting. We collected data on two plausible effort signals that teams receive from offensive linemen: penalties committed, which can reflect shirking and are monitored by a group of independent agents, the officials in NFL games, and sacks allowed, which are monitored only by the players and the coaches of the team. We find that sack yards allowed by other linemen on the team significantly reduce the earnings of offensive linemen, but sack yards given up by each lineman does not reduce his own earnings, other things equal. These results are consistent with the presence of a low power punishment mechanism through which a group of team members can punish another team member who shirks during their repeated workplace interaction. The results are also consistent with the presence of JPE wage schemes in this setting. Both are features of the
models developed by Che and Yoo (2001) and Ishida (2006), so the results here suggest that these models explain observed outcomes in the performance and earnings of NFL offensive linemen.

These results also extend the growing literature of the importance of peer effects in sport. Depken (2000) found evidence supporting peer effects in Major League Baseball. Guryan et al. (2009) found no evidence of peer effects in professional golf, based on a setting that includes random assignment of team mates. Depken and Haglund (2011) found evidence of peer effects in foot races. This analysis uses a novel setting, offensive line play in the NFL, which exploits the repeated nature of productive interaction in teams, and which generates plausible observable effort signals. Both these features are unique to the literature on peer effects in teams.

Finally, the presence of JPE wage schemes and viable weak powered enforcement mechanisms have important implications for research on the payroll-success relationship in professional sports. Many models in this literature assume that higher payrolls lead to increased production of wins; see, for example, the model developed by Fort and Quirk (1995). Che and Yoo (2001) show that JPE wage schemes generate incentives at a lower cost than RPE incentive schemes. If JPE wage incentive schemes are widespread in professional sports, then the relationship between total payroll and team success may not be as well-behaved or strong as many existing models of team production in sport assume, especially if some teams use JPE wage schemes and others use RPE schemes.

References


