Incentive Pay and Decision Quality: Evidence from NCAA Football Coaches

Marcella Cartledge* and Luke Taylor†

Abstract

Using play-by-play American football data and panel data on head coach remuneration, we test whether a head coach's incentive pay affects the quality of their decisions. We proceed by first estimating an 'optimal strategy' for first-down offensive plays, then investigate whether the gap between actual and optimal choices is affected by incentive pay. In contrast to merely looking at the outcome of an agent's choice, our approach considers the decision environment and the resources available. We find a small, but significant, negative effect of incentive pay on decision quality. Critically, this effect is not found when looking at raw outcome measures. (J33; C23)

Keywords and Phrases: Incentive Pay; Sports; Decision Quality; Semiparametric Estimation; Panel data

*University of Virginia, Department of Economics; Email: marcellacartledge@gmail.com; Address: Monroe Hall, Charlottesville, VA 22903, USA; Telephone: +1 434-924-3177.

†Corresponding author. Aarhus Universitet, Department of Economics and Business Economics; Email: lntaylor@econ.au.dk; Address: Fuglesangs Allé 4, 8210 Aarhus, Denmark; Telephone: +45 87 15 00 00. Financial support from the Aarhus University Research Fund (AUFF-26852) is gratefully acknowledged.

1 Introduction

Incentive pay is a popular tool used by governments and firms to elicit improved performance. It is used to motivate CEOs to improve the profitability of their firm (Cooper, Gulen and Rau, 2016), to incentivise sports players and coaches to perform at their best (Stiroh, 2007; Bryson, Buraimo and Simmons, 2011), to induce higher quality teaching in schools (Fryer, 2013), and to improve health care by doctors and nurses (Mullen, Frank and Rosenthal, 2010). However, despite its ubiquitous use, there is still conflicting evidence on the effect of incentive pay on performance.

In this paper, we use play-calling for offensive plays on first downs in NCAA[1] Division I American football as a lens through which to test whether pay-for-performance improves the quality of decision-making. We argue that in order to truly reflect the effect of incentive pay, it is not enough to look at the raw outcome of an individual's effort; we must benchmark this against some suitable level to account for the resources at their disposal. As our benchmark, we choose a 'model-optimal' outcome, which reflects the best potential outcome which could have been achieved in a given situation.

We abstract from the myriad of tasks a head coach is responsible for and concentrate only on the offensive strategy during matches. Using detailed play-by-play data on first downs and a nonparametric additive model with two-way fixed-effects (Mammen, Støve and Tjøstheim, 2009), we estimate a function to model the outcome of a given play. By maximising this function with respect to the choice of player to use in a given play, we construct a 'model-optimal strategy' given the abilities of the individual players, the strength of the opposition, and the current game situation. From this model, we create an efficiency statistic calculated as the difference between the outcome of the model-optimal choice and the actual choice. This enables us to measure a coach's performance conditional on the available resources - that is, the abilities of the players - and the task environment. With this measure, and with panel data on head coaches' remuneration contracts, we determine whether an increase in their maximum potential bonus improves relative performance.[2]

The key endogeneity issue arising in investigations of pay-for-performance is that individuals self-select into firms based on the level of incentive pay. If the most productive and highest quality individuals choose to work for firms with high levels of performance-related-pay, the effect of these incentive schemes will be biased upwards. To circumvent this problem, we use fixed-effects at the individual coach level.

Furthermore, to ensure the effect of incentive pay only works through the play-calling channel, we use head coaches who are *not* in charge of calling offensive plays as a control group; evidence is also provided for the suitability of this approach. We can then isolate the incentive pay effect through this channel by

[1] The National Collegiate Athletic Association (NCAA) is a non-profit organisation which serves as both the organiser and governing body of intercollegiate athletics in the United States.

[2] More than 99% of head coaches have some form of on-field performance related bonus stipulated in their contract.

looking at the difference between the effect for non-play-calling head coaches and play-calling head coaches, respectively.

The decision to focus only on offensive strategy allows us to simplify our analysis without compromising our results. American Football provides an unusually clean environment for testing the ability of a collective to optimise since the separation between a team's offence and their defence is absolute. That is, it provides a situation in which the maximisation of win probability is separable in offensive and defensive strategies. In almost all other team sports there is a complex interplay between attacking and defending that results in a trade-off between the two (see Hakes and Sauer, 2006; Berri, 2008; and Bradbury, 2008). Moreover, we heavily restrict the types of plays we consider such that we can be confident in the incentives and optimal strategies of a play-calling coach.

In general, the sports field has become an increasingly popular playground for economists to test their theories (see, for example, Simmons and Berri, 2011; Barros, Peypoch and Tainsky 2014; Lenten, 2017). Sports provide an ideal setting to study the behaviour of economic agents in a strategic environment. Similarly to auctions, the rules are well-defined and well understood, the payoffs are clear with potentially substantial incentives to acting rationally, there is ample opportunity to learn, and there is an abundance of rich data to analyse.

Furthermore, the sporting arena is particularly well-suited to studying the effects of incentives on performance. In a conventional work environment, it is difficult to determine an individual's output, especially when they are a small part of a large firm. Typically, it can also prove challenging to gauge the size of the incentive for a specific task since many employees perform numerous functions which are both varied and interrelated. Additionally, compensation is rarely linked to the outcome of an individual task. In contrast, in American football, there is a clear measure of output produced by a small number of people, the coach has a single job - to win games - and this represents the key measure on which bonuses are paid. However, the number of games won is not a measure that varies at the play level. As a result, our analysis is instead built on the popular 'expected points added' (EPA) measure. This captures the success of a single play while also ensuring that when it is summed over the entire game, the winning team will have the highest EPA.

Our main result shows that when measuring head coach performance in terms of our efficiency measure, i.e. benchmarked against the model-optimal choice, there is a small, but significant, negative effect of incentive pay on decision quality. In particular, a one standard deviation increase in maximum potential bonus leads to a 0.14 standard deviation decrease in our efficiency measure. However, and perhaps more importantly, when the performance measure used is the raw outcome (not relative to the model-optimal benchmark), the effect on incentive pay drops by 50%, to 0.07 of a standard deviation, and is insignificant.

This result adds weight to the ability of our benchmarking method to take into account the resources available. In any analysis of the effect of incentive pay, an individual's performance rarely occurs in a vacuum. Although in our context, the environment relates to the play situation and the ability of the players, in any other situation, a similar set of environmental factors will affect the outcome of an individual's actions.

The rest of this paper proceeds as follows. Section 2 gives a brief summary of the related literature. In Section 3, we provide details of the model we use to answer this question. Section 3.4 outlines the estimation procedure and contains an explanation of the data. Section 4 describes and explains our results. Finally, Section 5 concludes.

2 Related Literature

In contrast to the psychology literature, the economics literature has found ample evidence supporting the use of pay-for-performance schemes. Lazear (2000), for example, uses data from Safelite Glass Corporation to show that an incentive payment scheme led to higher firm productivity equivalent to a 44% increase in output per worker. While Lavy (2009) shows that monetary incentives for math teachers in Israel improved test-taking rates among students, as well as mean test scores among students.

In the psychology literature, however, there is a large body of experimental and field research that refutes the claims above. Camerer, Lowenstein and Prelec (2005) use MRI data to show that risky decisions that result in gains and losses are coded separately in the brain in a way that implies agents experience hyperbolic discounting. This challenges the economic assumption that monetary incentives would necessarily lead to improved decision-making in risky situations. Ariely et al. (2009) conduct a series of experiments in which they vary the size of the monetary reward tied to performance on a given task. They discuss several reasons why increased incentives may even lead to a deterioration in performance, namely: the agent may be too focussed on the reward rather than the task, there may be a shift from 'automatic' to 'controlled' mental processing, or because of the 'Yerkes–Dodson law', which states there is an optimal level of arousal for executing tasks, and any mental engagement above or below that level will result in suboptimal performance (see also Savage and Torgler, 2012).

Frey and Jegen (2001) discuss another reason why incentive pay may decrease performance: the "motivation crowding effect". Providing subjects with extrinsic monetary incentives may undermine intrinsic motivation to complete a task well. McGraw and McCullers (1979) find experimental evidence consistent with this theory; they show that subjects' performance decreased on a simple task relative to performance on a more complicated task when monetary incentives were introduced.

Finally, Baumeister (1984) provides experimental evidence that decreased performance may be a result of subjects "choking under pressure", defined as worsened performance under circumstances in which a good performance is especially important. Yu (2015) uses neuroimaging data to reveal that the area of the brain responsible for motivation also plays a key role in stress-induced choking; and in high-stakes situations, this part of the brain may be over-stimulated, leading to poor performance. Interestingly, Bühren and Gabriel (2021) find contrasting evidence in this area. In their recent paper, they find that handball players' penalty performance improves when the thrower's team is behind, suggesting that increased stakes can actually improve performance.

3 Empirical Strategy

The approach we use contains several steps. For clarity, we briefly outline each step without going into detail, then take each part in turn. First, we build a model to describe the 'outcome' of a single play based on the 'strategy' used. In the second step, we use this model to predict the outcome of each strategy that could have potentially been used in a given play. Next, we create a measure of efficiency using the fitted value from the chosen strategy and the fitted value from the 'model-optimal strategy'. Finally, we determine how this efficiency measure is related to the maximum potential bonus of the head coach in charge of that play.

3.1 Model

The head coach is an economic agent who makes decisions to maximise their utility. In general, incentive pay contracts for coaches are based on the number of wins the team achieves. To increase the probability of winning, an American football offence scores points. Of course, the performance of the defence in preventing points being scored will also affect win probability and hence enter the utility function of the coach. However, as discussed in Section 1, we can confine our analysis to offensive play situations because win probability is separable in a given team's offensive and defensive strategies. Hence, the coach's offensive maximisation problem is not affected by the choice of defensive strategy.

We assume an offensive strategy is characterised by the player chosen to advance the ball given the current situation and the preceding plays. This may either be through the quarterback passing the ball to a receiver or a runner carrying the ball. The rules of American football lead to a strategic setting where the optimal strategy depends heavily on the current state of play. Each team is given four chances, or 'downs', to advance the ball towards the goal line but, crucially, if you move forward ten yards or more, your four chances are reset (in football terminology, 'you make a first down'). Although seemingly harmless, the resetting of one's chances affects the optimal strategy considerably. Without this additional rule, the optimal tactic would be to maximise expected yards *for a single play,* and the optimisation would be straightforward.

However, the team must instead consider how their strategy affects the variance of expected yards in a single play - since this impacts the likelihood of making a first down and hence their overall number of chances. To circumvent this complex task of balancing the expected number of yards and their variance, we do not consider optimising a function of yards, but instead, maximise a popular measure of play-outcome known as 'expected points added' (EPA).

The aim of a maximising coach is to win the game. However, this is not a measure that varies at the play level. EPA was first introduced by Carter and Machol (1971) as a measure that captures the success of a single play while also ensuring that when it is summed over the entire game, the winning team will have the highest EPA. There are several different approaches to calculating EPA, but all are based on the same underlying idea; we outline only the approach we use. The first step is to determine the expected number of points a team will make in a drive[3] when starting from a given game situation. We define a game situation by the down that the offensive team is on, the distance until the next first down, and the field position. To calculate this expectation, we use a third-degree polynomial regression with points scored at the end of the drive as the dependent variable and the game situation variables previously described as predictors. We can then determine the change in expected points from a single play, i.e. EPA, as the difference between the expected points for the game situation at the start of the play and the expected points for the game situation at the end of the play.

By using EPA, we align our mathematical model as closely as possible to the implied model used by the head coach. To further ensure this alignment, we restrict the situations for which we calculate EPA and for which we attempt to find an optimal strategy. Our goal is to prevent any possible sources of divergence between the optimisation problem we solve, and that solved by the coach.

Firstly, we consider only first down plays with ten yards to go until the next first down. Strategies will be different depending on the down and the distance until the next first down. By considering only one scenario, we need only consider one strategy. Furthermore, when we come to estimate player abilities (as fixed-effects from a generalised additive panel model), we want to ensure we consider similar plays on the first down. We chose this particular play type as it is the most frequently encountered; this gives us the greatest number of observations and ensures that the coach has experienced many plays of this type.

Secondly, we confine ourselves to plays in the first quarter, the first half of the second quarter, and the third quarter. This removes any possibility for changes in strategy or player abilities due to nearing the end of the game (in the fourth quarter), or from nearing the end of the half in the second quarter (the third quarter restarts from the centre line and the offence may have to give up possession of the ball).

Thirdly, we only use plays for which the absolute difference in the score is less than 14 points. This acts

[3] A drive is the terminology used for a series of plays for the offence before possession is given to the opposing team.

to prevent any changes in strategy or player abilities as a result of being very far behind, or very far in front of, the opposition; each team can still win or lose the game with two converted touchdowns.

Finally, we remove plays which start within 40 yards from the opponent's goal line or start within 20 yards from the offensive team's goal line. The first restriction ensures that it is never sensible to attempt a field goal. It is commonly believed that beyond the 35-yard line a field goal is improbable; together with the restrictions on the quarter of play and the play situation, no field goal attempts are made in our selected sample. It also prevents any changes in strategy as the offence approaches the goal line. The decision to remove plays within 20 yards of one's goal line is to prevent any changes in strategy resulting from caution at being so close to one's goal line.

By imposing these restrictions, we can remain agnostic about how the strategy could be affected by these variables. The narrowness of the decision environment should not be confused with the narrowness of the question. The aim of restricting the task so severely is to ensure that the model we solve is the same as the model solved by the agents. We must find a compromise between the clean - but potentially unrealistic - setting of a laboratory, and the messy real world. Our approach is to focus on a single, real-world task by removing as many potential concerns as possible. As a final protection against these additional factors affecting the strategy, in our regression results in Section 4, we control for the position on the field, the points difference, and the quarter of play (each of which is insignificant in our models).

Offensive tactics can be broadly split into passing to a receiver or allowing a running back to carry the ball - known as 'passing' and 'rushing', respectively. How each approach generates EPA is likely to be different; not least because the opposition defence may be more adept at defending one type of attack. As such, we estimate separate maximisation functions for each type of play.

For each maximisation function, we assume that only two individuals (or groups of individuals) are involved. In the case of passing plays, a receiver works with his quarterback to produce EPA; for rushing plays, a running back works with his offensive line. The passing play function is given by

$$Y_{ijpg} = \alpha_i + \beta_j + \sum_{k=1}^{\kappa} m_k \left(X_{ijpg}^{(k)} \right) + \epsilon_{ijpg}$$

where Y_{ijpg} represents the EPA produced by receiver i with quarterback j in play p of game g, α_i represents the ability of receiver i, β_j represents the ability of quarterback j, ϵ_{ijpg} represents the regression error, $m_k(\cdot)$ are nonparametric functions to be estimated, and $X_{ijpg}^{(k)}$ are a set of regressors including the defensive quality of the opposing team (described in Section 3.3), the number of previous attempts for the receiver i and quarterback j pair in game g, the number of previous plays in the current drive, and the field position.

An analogous maximisation function for rushing EPA is given as

$$Y'_{i'j'pg} = \eta_{i'} + \lambda_{j'} + \sum_{k=1}^{\kappa} m_k \left(X^{(k)}_{i'j'pg} \right) + v_{i'j'pg}$$

where $Y'_{i'j'pg}$ represents the EPA produced by running back i' with offensive line j' in play p of game g, $\eta_{i'}$ and $\lambda_{j'}$ represent the ability of running back i' and offensive line j', respectively. The same regressors are used as in the passing function and all other parameters and variables are defined analogously.

3.2 Using the Model

For each play, and for each player available in that play,[4] we use the relevant estimated maximisation function to predict the EPA that would have been produced if the ball had been given to that available player. Importantly, we take all strategies from past plays as given. The player who is predicted to produce the greatest EPA in a given play is deemed the 'model-optimal choice'. For each play, we construct a measure of efficiency defined as the difference between the predicted optimal EPA and the fitted value for the strategy chosen by the head coach. By using the fitted value of EPA for the actual choice rather than the observed EPA, we account for the idiosyncratic performance of the players. That is, we compare the conditional expectation of EPA from the optimal choice with the conditional expectation of EPA from the actual choice.

The final step is to determine the effect of the head coach's maximum potential bonus on the efficiency of their decisions. To do this, we merge our play-by-play data, including our efficiency measure, with panel data on head coach remuneration contracts and estimate a regression with head coach fixed-effects.

With panel data at the individual coach level, we hope to alleviate fears of endogeneity. Including coach fixed-effects allows us to control for, in particular, both the innate ability of the head coach and their preferences for a particular incentive pay structure. The most pressing concern with any analysis of pay-for-performance is that individuals sort themselves into firms with a compensation scheme that matches their preferences. For example, efficient workers move to firms with high incentive pay schemes, in doing so, they inflate the true effect of incentive pay on performance; fixed-effects at the coach level will remove such concerns.

Furthermore, to ensure the effect of incentive pay only works through the play-calling channel, we use head coaches who are *not* in charge of calling offensive plays as a control group. This group of non-play-calling head coaches make up 71% of all head coaches in our sample. We can then isolate the incentive pay effect through this channel by looking at the difference between the effect for non-play-calling head coaches

[4]We are only able to consider players as being available if they are playing in the given game and also appear in our restricted sample at some point (not necessarily in that particular game).

and play-calling head coaches, respectively.

3.3 Data

We use play-by-play data for seven seasons, 2007 - 2013,[5] for every game in the NCAA Division I American football league to estimate our two maximisation functions. As mentioned in Section 3, we restrict our sample to first down plays with 10 yards until the next first down. We also only consider plays in the first quarter, the first half of the second quarter, or the third quarter, and plays where the points difference is less than 14. Finally, we remove all plays which start within 40 yards of the opponent's goal line or 20 yards of the offensive team's goal line. These data restrictions are made to ensure a degree of homogeneity in the observations and to hone in on the effect of interest.

For our defensive ability measures, we use leave-one-out averages. Specifically, for our measure of defensive ability against passing plays, for each team in each season, we take the average EPA for all passing plays (except the play in question) against that defensive team. We use an analogous measure for rushing plays.

To determine the effect of the head coach's maximum potential bonus and their decision quality, we merge this play-by-play data with panel data for each head coach's remuneration over the seven seasons. This remuneration data is obtained from USA Today who made freedom of information requests to all schools for all forms of compensation for head and assistant coaches. Twenty schools are private, or they are public schools where state law exempts the release of coach salary information, and so are removed from the analysis. 'School Pay' is defined as any pay (even if paid by another source, such as an apparel company) that is guaranteed by the school, including base salary and other income from contract provisions such as deferred payments and one-time bonuses (e.g. a signing bonus). It does not include amounts that might have been earned as annual incentive bonuses in previous years, or the value of standard school benefits, such as health care. The 'Maximum Bonus' variable refers to the highest amount the coach can receive if the team meets on-field performance goals (predominantly win totals but this may also include bowl-game appearances and championships), academic goals or player conduct goals. It does not include payments based on ticket revenue or sales.

We also manually collected data through online coaching profiles on the tenure for each coach and whether they were in charge of calling offensive plays. The tenure variables include total time spent as a head coach for an NCAA or NFL team, and total consecutive time spent at their current team.

After merging these datasets and restricting the sample, we have 22 866 observations in the baseline model. Descriptive statistics for all variables are given in Tables 1 - 3 below.

[5] We use data only up to 2013 since we do not have remuneration data past this point.

Table 1: Descriptive Statistics (Rushing Maximisation Function)

Statistic	N	Mean	St. Dev.	Min	Max
EPA	43,547	−0.04	0.66	−2.55	4.22
Defensive Ability - Rush	43,547	0.05	0.11	−0.37	0.41
Attempt	43,547	5.48	4.98	1	37
Play Number (in Drive)	43,547	4.43	4.79	1	50
Field Position	43,547	61.57	10.85	41	79

Table 2: Descriptive Statistics (Passing Maximisation Function)

Statistic	N	Mean	St. Dev.	Min	Max
EPA	28,976	0.22	0.86	−2.17	4.22
Defensive Ability - Pass	28,976	−0.21	0.15	−0.94	0.60
Attempt	28,976	2.88	2.34	1	19
Play Number (in Drive)	28,976	4.52	4.81	1	45
Field Position	28,976	61.27	10.78	41	79

Table 3: Descriptive Statistics (Regression Model)

Statistic	N	Mean	St. Dev.	Min	Max
Efficiency	22,866	−0.70	0.48	-2.18	0.00
EPA	22,866	0.06	0.77	−2.42	4.22
Play Caller × Maximum Bonus	22,866	185,034.60	382,753.70	0	1,800,000
Maximum Bonus	22,866	604,373.30	445,407.20	0	3,159,000
Play Caller	22,866	0.29	0.45	0	1
Pay	22,866	1,805,546.00	1,219,809.00	190,000	5,997,349
Assistant Pay	22,866	1,883,488.00	875,435.20	262,279	4,565,803
Assistant Maximum Bonus	22,866	319,984.50	397,438.20	0	2,621,500
Play Number	22,866	4.57	4.93	1	50
Attendance	22,866	50,248.01	26,286.03	1,349	115,109
Home	22,866	0.45	0.50	0	1
Field Position	22,866	61.84	10.93	41	79
Points Difference	22,866	−0.78	5.90	−13	13
Winning	22,866	0.30	0.46	0	1
Options Variance	22,839	0.18	0.13	0	1.59
Options Number	22,866	9.20	2.04	2	16
Defensive Ability (Pass)	22,866	−0.20	0.16	−0.85	0.32
Defensive Ability (Rush)	22,866	0.06	0.11	−0.30	0.39
Tenure (Team)	22,866	4.67	5.58	0	26
Tenure (Total)	22,866	8.70	8.20	0	36
Transfer	22,866	0.17	0.38	0	1
Game Number	22,866	5.04	2.85	1	14

3.4 Estimation

In the first stage of our analysis, we are only concerned with obtaining the best possible predictor of the two maximisation functions. We do not need consistent estimates of an underlying regression function nor do we make any claims of causality. Unfortunately, suitable machine learning procedures designed for two-way panel data settings are lacking. However, we can use an extension of the generalised additive model, first put forward by Hastie and Tibshirani (1990), which allows for additive fixed-effects. In particular, we use the smoothed backfitting approach of Mammen, Støve and Tjøstheim (2009). For completeness, and because our situation is slightly more complicated than that presented in Mammen, Støve and Tjøstheim (2009), we provide details of their estimator in Appendix A.1. Here, we simply outline the structure of our data.

As well as the panel being unbalanced, each receiver often only ever plays with one quarterback (likewise for running backs playing with offensive lines). Thus, the receivers can be viewed as repeated cross-sections over different quarterbacks and not a true panel. The within transformation is routinely used to remove individual fixed effects, i.e. in a conventional panel data model, subtract $\frac{1}{T}\sum_{t=1}^{T} Y_{it} = \bar{Y}_i$. However, with a repeated cross-section we only observe each individual once so would not be able to take an average over time. Constructing cohorts or using instrumental variable procedures have been proposed to deal with this issue. Fortunately, we have the luxury of four panel-dimensions in our data while only using two-way fixed-effects. Thus, supposing we used a simple linear model, the receiver fixed-effects could be removed by subtracting

$$\frac{1}{n_{Re(i)}} \sum_{g=1} \sum_{j=1} \sum_{p=1} Y_{ijpg} = \bar{Y}_i,$$

where $n_{Re(i)}$ denotes the total number of observations on receiver i (quarterback fixed-effects being obtained analogously). This same reasoning is easily extended to the generalised additive model, as shown in Appendix A.1.

4 Results

4.1 Main Results

We first discuss the performance of our generalised additive panel model in describing the relationship between EPA and our predictors, namely: the quality of the defensive team, the number of previous attempts for the chosen player, the number of previous plays in the current drive, and the field position. There is a slightly better fit for the regression explaining passing plays in comparison to rushing plays. The in-sample correlation between actual EPA and estimated EPA for the passing play function is 0.46; for rushing plays, the correlation is 0.40.

Figure 2 in Appendix A.1 plots the distributions of the player fixed-effects obtained from the generalised additive panel model for receivers, running backs, quarterbacks, and offensive lines, respectively, which we interpret as the ability of each respective player. There are some differences in these distributions which can be thought of as differences in the distributions of abilities. The standard deviations for receivers, running backs, quarterbacks, and offensive lines are 0.46, 0.37, 0.34, and 0.32, respectively. Unsurprisingly, offensive lines display the smallest standard deviation since an offensive line is made up of several players, and is, therefore, an average of their abilities. It is also unsurprising that the standard deviation of quarterback abilities is lower than that of running backs or receivers. This is because, typically, only two or three quarterbacks will be used by a team in any given season, in comparison to more than ten receivers (or running backs); the variance of the maximum of a set of random variables is always smaller than the variance of the largest $n > 1$ of those random variables.

Our baseline model is given in Table 4. The regressor of interest is the interaction between the maximum potential bonus of the head coach (Maximum Bonus) and the dummy variable indicating whether the head coach is in control of the offensive play strategy (Play Caller). We control for the head coach's base salary (Pay), the average base salary and maximum potential bonus of the assistant coaches[6] (Assistant Pay and Assistant Maximum Bonus, respectively), the number of previous plays in the current game (Play Number), the size of the crowd at the game (Attendance), whether the offensive team is playing at home (Home), the field position at the start of the current play (Field Position), the difference in points at the start of the current play (Points Difference), whether the offensive team is currently winning (Winning), the variance of the fitted values of the potential players in the current play (Options Variance), and the number of potential players in the current play (Options Number).

The left column uses our efficiency measure as the dependent variable, calculated as the difference between the fitted value of the 'model-optimal choice' and the fitted value of the actual choice. A Tobit estimator is used to account for the fact that efficiency is bounded from above by 0. The middle column uses OLS with the raw EPA measure as the dependent variable. Finally, the right column uses a probit estimator with a binary variable for whether the team wins the game as the dependent variable; the marginal effects are reported in the table. To calculate the marginal effect for the interaction term in the Tobit and probit models, we follow Ai and Norton (2003).

In all three regressions, we also add coach, season and conference fixed effects, quarter of the game, game number of the current season, total tenure of the head coach, tenure at the current club, whether the coach has transferred to the team in this season (as this can affect their remuneration packet), and ability of the

[6]This helps capture the wealth of the team and controls for its impact on performance through assistant coaches.

defensive team to defend against passing plays and rushing plays, respectively. Standard errors are given in parentheses; we use heteroskedasticity robust pseudo-jackknife "HC3" errors (MacKinnon and White, 1985) clustered at the head coach level (Cameron, Gelbach, & Miller, 2011). All continuous variables (including the dependent variables) are standardised to have unit variance.

The key result of Table 4 is the significant negative effect of incentive pay on the quality of the play strategy captured by the coefficient on the interaction between bonus and play-caller. In particular, this coefficient indicates that the effect of a one standard deviation increase in the maximum potential bonus on efficiency for a play-calling coach is 0.14 of a standard deviation lower than for a non-play-calling coach. In other words, a higher maximum potential bonus reduces the ability of a coach to make good decisions in the game. Critically, this effect is twice that found using raw EPA, and more than four times the magnitude of the effect using the binary variable for winning the game.

According to the results using the efficiency measure as the dependent variable, increasing the maximum potential bonus of a non-play-calling coach from the 25th percentile (\$265 000) to the 75th percentile (\$830 000) leads to an increase of 0.02 EPA (holding the optimal achievable EPA constant). This is in comparison to a 0.07 decrease for a play-calling coach - giving a 0.09 drop in EPA (or a 0.5 percentage point decrease in win probability) per play attributable to decision making. It is tempted to aggregate these findings to the game level, rather than a single play. However, our analysis uses just a subset of all plays (and does not consider defensive plays at all); thus, we should be cautious extrapolating our results beyond what they are able to show.

In contrast, again moving from the 25th percentile to the 75th percentile of maximum bonus, the results using raw EPA show a 0.04 EPA increase for a non-play-calling coach, and a 0.001 decrease for a play-calling coach - giving a 0.04 drop in EPA (or a 0.2 percentage point decrease in win probability) per play attributable to decision making. Finally, for the binary outcome case, there is a 1.2 percentage point increase in win probability for a non-play-calling coach and a 0.9 percentage point increase in win probability for a play-calling coach - resulting in an overall 0.3 percentage point decrease per play attributable to decision making. Figure 1 below gives a graphical representation of these results.

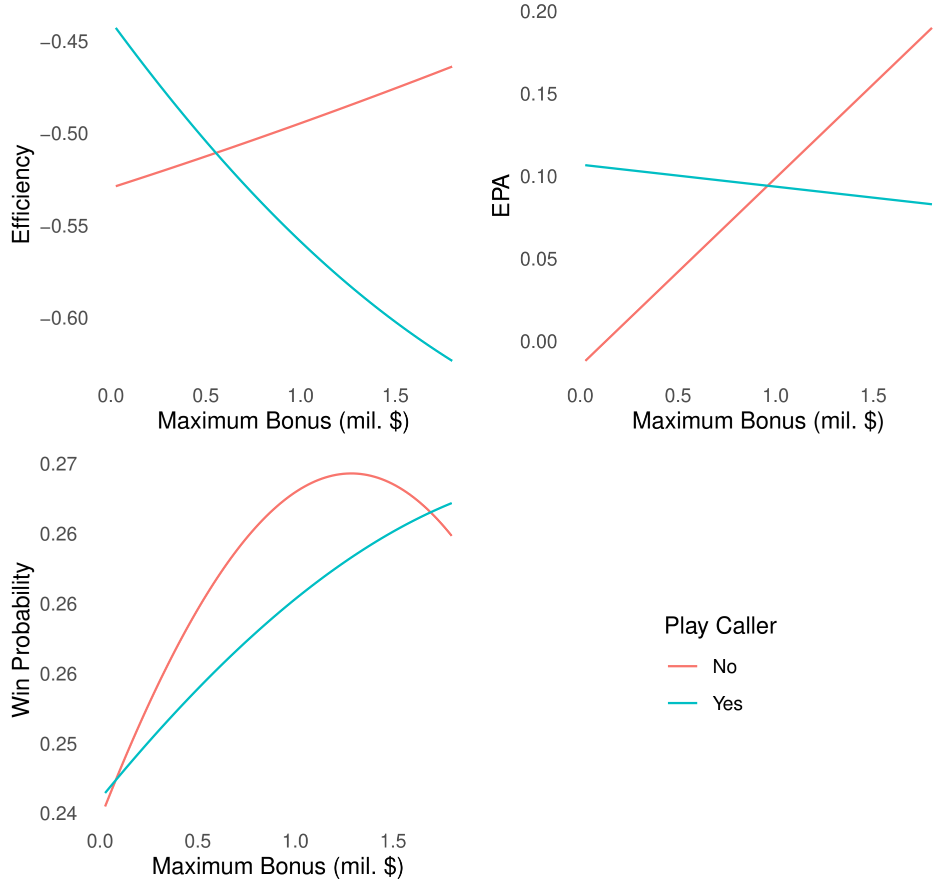
While the effects of the maximum potential bonus are relatively small, the differences in these results across the three methods indicate the importance of our approach. If one is interested in determining the effect of incentives on performance, it is not enough to merely look at outcomes. We must find a convincing way to assess performance relative to a benchmark. In most cases, it makes sense to set this benchmark as the optimal performance, as we do in this paper; however, this need not be the only choice.

Finally, it is worthwhile comparing the coefficient on the interaction term with the coefficient for the bonus term itself. Specifically, the value of 0.02 for the maximum bonus variable captures the effect of incentive pay

Table 4: Regression Results (Marginal Effects)

	Dependent variable:		
	Efficiency	EPA	Win
	(Tobit)	*(OLS)*	*(Probit)*
Maximum Bonus × Play Caller	-0.14^{***}	-0.07	-0.03^{*}
	(0.02)	(0.04)	(0.03)
Maximum Bonus	0.02^{*}	0.07^{**}	0.09^{***}
	(0.01)	(0.03)	(0.02)
Play Caller	0.05	0.06	-0.05
	(0.05)	(0.10)	(0.06)
Pay	-0.01	-0.09^{***}	0.07^{***}
	(0.01)	(0.03)	(0.02)
Assistant Pay	0.04^{**}	0.07^{**}	-0.05^{***}
	(0.01)	(0.03)	(0.02)
Assistant Maximum Bonus	0.01	-0.02	0.01
	(0.01)	(0.02)	(0.01)
Play Number	0.002	0.01^{*}	-0.02^{***}
	(0.003)	(0.01)	(0.004)
Attendance	0.01^{**}	0.04^{***}	0.09^{***}
	(0.005)	(0.01)	(0.01)
Home	-0.01^{*}	0.06^{***}	0.11^{***}
	(0.01)	(0.01)	(0.01)
Field Position	0.001	0.002	-0.02^{***}
	(0.003)	(0.01)	(0.01)
Points Difference	-0.01	-0.01	0.24^{***}
	(0.01)	(0.01)	(0.01)
Winning	0.02^{*}	0.06^{***}	0.23^{***}
	(0.01)	(0.02)	(0.01)
Options Variance	-0.31^{***}	0.03^{***}	-0.02^{***}
	(0.003)	(0.01)	(0.004)
Options Number	-0.07^{***}	-0.01	0.02^{***}
	(0.03)	(0.01)	(0.004)
Log Sigma	-0.20^{***}		
	(0.01)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 1: Effect of Maximum Bonus on Outcomes



outside of the in-game offensive strategy, for example, work carried out during training sessions. Whereas, the coefficient associated with the interaction term captures the difference in the effect of incentive pay between coaches who control offensive play strategy and those who do not, i.e. it captures the effect through the play-calling channel. We discuss the robustness of this result in Section 4.2.

4.2 Alternative Explanations and Robustness Checks

It is essential to consider any other potential avenues which could explain our findings. First, it could reasonably be argued that, while still attempting to maximise EPA, coaches are likely to want to 'test out' new players or to give some players an attempt so that morale does not suffer. To check the robustness of our results to such situations, we remove all plays involving a player who has been used less than five times (set as the 25th percentile of player usage). Our results are displayed in Table 5 in Appendix A.2 and show a very similar story to that given in Table 4.

It is encouraging to note that neither the field position nor the points difference has a significant impact on efficiency. This suggests that we have restricted our sample sufficiently well such that coaches do not follow different strategies according to these variables. If they did, we should see a change in efficiency under different play settings because they are solving different problems. Of course, this does not necessarily prove that the coaches are solving the same optimisation as us.

Another concern in the alignment of our model to that of the coaches relates to the timing of plays during a game. Consider the following scenario. A head coach has access to a star player, but this player has low stamina and tires quickly. In response to this, the head coach may decide not to use this player until the end of the game. Whether this is an optimal strategy in itself is grounds for debate. However, we cannot prove that this is an incorrect strategy. To investigate whether this is a valid concern, we include in our analysis the play number within the current game. If it is true that the coach keeps certain players for specific times in the game, irrespective of the current game situation, we should see efficiency change based on the play number. For example, in the case with the 'low-stamina star player' left to the end of the game, we would see that efficiency increases as we move later in the game. However, we can see that this is not the case: play number has a small and insignificant effect on efficiency.

In order for our method of using non-play-calling head coaches as a control group to isolate the effect of incentive pay on decision quality, we require that the two types of coaches, play-callers versus non-play-callers, are equivalent conditional on our set of covariates. To give weight to this assumption, we appeal to the coefficient on the dummy variable for play-calling responsibility displayed in Table 4. The insignificant effect seen in each regression indicates that whether or not the head coach calls plays has little impact on the efficiency measure, the EPA, or the win probability. Furthermore, a linear regression of maximum potential bonus on play-calling status (including all other control variables and with standard errors clustered at the head coach level) results in an insignificant effect of 0.06 with a p-value of 0.77; see Table 6 in Appendix A.2. These findings provide strong evidence to the similarity of the two types of coaches and hence the validity of our main results.

Finally, despite the structure of the pay contract seeming to have little effect on performance, this does not necessarily mean that the contract has no effect. As a result of using individual coach fixed-effects, the results obtained are unaffected by coaches of different quality sorting based on remuneration. It may still be that paying high base salaries, or high bonuses, can attract the best coaches. However, when we estimate the regressions from Table 4 without coach fixed-effects (shown in Table 7 in Appendix A.2) we see that the size of the bonus has no effect on decision quality nor raw EPA. Moreover, the base salary is significantly negatively related to each dependent variable. It appears that the highest quality head coaches are not 'following the money'.

4.3 Discussion

In relation to the previous literature, our result of a small negative effect seems to agree more closely with the research carried out in psychology rather than economics. This is not to say that the economics literature shows unambiguous positive results, indeed, several prominent papers also find a detrimental effect of incentive pay on performance (Ariely et al., 2009; and Fryer, 2013). However, the focus of this paper is to highlight the need to account for the resources at the agent's disposal and the environment in which the agent is making choices. In the setting used in this paper, it is quite conceivable that the teams which pay the most in bonuses are also the teams which have the best players. This would lead to an overestimate of the effect of incentive pay on performance.

Our results have implications for the design of remuneration contracts for highly skilled employees. Neither the effect of the incentive pay nor the effect of the base salary is economically substantial, indicating that the type of pay structure has little ability to affect the performance of the coach. This may be because the individual is already exerting their maximum potential effort, or they are motivated by performing well and not directly by money.

The biggest concern when attempting to use these results to guide contract design relates to external validity. There is an inherent trade-off between restricting the environment under analysis to align the theoretical model with reality, and external validity. By considering only first down plays in the least consequential parts of the game (according to field position and time left to play), we limit potential discrepancies between the optimisation problem solved by the coach and that solved by us. However, this raises the question of whether our findings are specific to this narrow task environment or if they can be extrapolated to other highly skilled tasks. We believe that our setting provides a reasonable compromise between the artificial confines of a controlled laboratory experiment - which is unlikely to replicate the high stakes of many real-life tasks and, consequently, its findings cannot be applied to non-laboratory scenarios - and the messy reality of unrestricted observational data.

5 Conclusion

In this paper, we estimate a 'model-optimal strategy' for offensive first down plays in NCAA American football using detailed play-by-play data and an additive nonparametric two-way fixed-effects estimator. With this model, we create an efficiency statistic which enables us to measure the quality of coaches' decisions subject to their resource and environmental constraints. Together with panel data on head coach remuneration packages and using head coaches who are not in charge of offensive plays as a control group, we show that if our efficiency statistic is used as the outcome variable, there is a small but significant negative effect of incentive pay on performance. In contrast, when raw performance outcomes are used to measure performance, we find smaller insignificant effects. The discrepancy between these results highlights the importance of accounting for the task environment and resources available by comparing to a relevant benchmark.

A Appendix

A.1 Estimation Procedure

To reduce notation, we explain only the estimation of the passing function; the rushing function is estimated analogously. We use n to denote the total sample size, while the total number of units in each dimension are given as $1 \leq i \leq n^{Re}$, $1 \leq j \leq n^{QB}$, $1 \leq p \leq n^{P}$ and $1 \leq g \leq n^{G}$. We define $n_{QB(j)}$ to be the total number of observations on quarterback j, $n_{Re(i)}$ to denote the total number of observations on receiver i, and so on.

Define $K_b(\cdot) \equiv \frac{1}{b}K\left(\frac{\cdot}{b}\right)$ where b is a bandwidth parameter and $K(\cdot)$ is a suitably chosen kernel function (see Mammen, Støve and Tjøstheim, 2009, for full details). Also define the following density estimators

$$
\begin{aligned}
\hat{f}_k(x) &= \frac{1}{n}\sum_{ijpg} K_b\left(X_{ijpg}^{(k)} - x\right) \\
\hat{f}_k^j(x) &= \frac{1}{n_{QB(j)}}\sum_{ipg} K_b\left(X_{ijpg}^{(k)} - x\right) \\
\hat{f}_k^i(x) &= \frac{1}{n_{Re(i)}}\sum_{jpg} K_b\left(X_{ijpg}^{(k)} - x\right) \\
\hat{f}_{k,h}(x_1, x_2) &= \frac{1}{n}\sum_{ijpg} K_b\left(X_{ijpg}^{(k)} - x\right) K_b\left(X_{ijpg}^{(h)} - x\right).
\end{aligned}
$$

The estimators of $m_k(\cdot)$, α_i, and β_j are defined as the minimisers of the following smoothed least-squares loss function

$$
\sum_{ijpg} \int \cdots \int \left(Y_{ijpg} - \hat{\alpha}_i - \hat{\beta}_j - \sum_{k=1}^{\kappa} \hat{m}_k(u_k)\right)^2 \prod_{h=1}^{\kappa} K_{b_h}\left(X_{ijpg}^{(h)} - u_h\right) du_1 \dots du_\kappa
$$

subject to

$$
\begin{aligned}
\int \hat{m}_k(u)\, \hat{f}_k(u) du &= 0, \text{ for } k = 1, \dots, \kappa \\
\sum_{i=1}^{n^{Re}} \frac{n_{Re(i)}}{n} \hat{\alpha}_i &= 0.
\end{aligned}
$$

This results in the following closed-form estimators

$$
\begin{aligned}
\hat{m}_k(x) &= \tilde{m}_k(x) - \sum_{i=1}^{n^{Re}} \frac{n_{Re(i)}}{n} \hat{\alpha}_i \frac{\hat{f}_k^i(x)}{\hat{f}_k(x)} - \sum_{j=1}^{n^{QB}} \frac{n_{QB(j)}}{n} \hat{\beta}_j \frac{\hat{f}_k^j(x)}{\hat{f}_k(x)} - \sum_{h \neq k} \int \hat{m}_h(u) \frac{\hat{f}_{k,h}(x,u)}{\hat{f}_k(x)} du, \text{ for } k = 1, \dots, \kappa \\
\hat{\alpha}_i &= \tilde{\alpha}_i - \sum_{j=1}^{n^{QB}} \frac{n_{Re(i),QB(j)}}{n_{Re(i)}} \hat{\beta}_j - \sum_{k=1}^{3} \int \hat{m}_k(u)\, \hat{f}_k^i(u) du, \text{ for } i = 1, \dots, n^{Re}
\end{aligned}
$$

$$\hat{\beta}_j = \tilde{\beta}_j - \sum_{i=1}^{n^{Re}} \frac{n_{Re(i),QB(j)}}{n_{QB(j)}} \hat{\alpha}_i - \sum_{k=1}^{3} \int \hat{m}_k(u) \hat{f}_k^j(u) du, \text{ for } j = 1, \ldots, n^{QB}$$

where $n_{Re(i),QB(j)}$ denotes the total number of observations involving both receiver i and quarterback j, and

$$\begin{aligned}
\tilde{m}_k(x) &= \frac{1}{n} \sum_{ijpg} K_{b_k}\left(X_{ijpg}^{(k)} - x\right) Y_{ijpg} / \hat{f}_k(x) \\
\tilde{\alpha}_i &= \frac{1}{n_{Re(i)}} \sum_{jpg} Y_{ijpg} \\
\tilde{\beta}_j &= \frac{1}{n_{QB(j)}} \sum_{ipg} Y_{ijpg}.
\end{aligned}$$

An iterative procedure based on these equations and the constraints can be used to obtain the estimates. We denote $\hat{m}_k^{[0]}(\cdot)$, $\hat{\alpha}_i^{[0]}(\cdot)$, and $\hat{\beta}_j^{[0]}(\cdot)$ as the starting values. For a given point of interest $x = (x_1, \ldots, x_\kappa)$, where $x_1, \ldots, x_\kappa$ may be vectors,

1. Set $a = 0$ and take $\hat{m}_k^{[a]}(x_k) = \tilde{m}_k(x_k)$, $\hat{\alpha}_i^{[a]} = \tilde{\alpha}_i$, and $\hat{\beta}_j^{[a]} = \tilde{\beta}_j$.

2. For $k = 1, \ldots, \kappa$ calculate

$$\hat{\hat{m}}_k^{[a+1]}(x_k) = \tilde{m}_k(x_k) - \sum_{i=1}^{n^{Re}} \frac{n_{Re(i)}}{n} \hat{\alpha}_i^{[a]} \frac{\hat{f}_k^i(x_k)}{\hat{f}_k(x_k)} - \sum_{j=1}^{n^{QB}} \frac{n_{QB(j)}}{n} \hat{\beta}_j^{[a]} \frac{\hat{f}_k^j(x_k)}{\hat{f}_k(x_k)} - \sum_{h \neq k} \int \hat{m}_h^{[a]}(u) \frac{\hat{f}_{k,h}(x_k, u)}{\hat{f}_k(x_k)} du.$$

3. Impose the constraint for $k = 1, \ldots, \kappa$

$$\hat{m}_k^{[a+1]}(x_k) = \hat{\hat{m}}_k^{[a+1]}(x_k) - \int \hat{\hat{m}}_k^{[a+1]}(x_k) \hat{f}_k(u) du.$$

4. For $i = 1, \ldots, n^{Re}$, calculate

$$\hat{\hat{\alpha}}_i^{[a+1]} = \tilde{\alpha}_i - \sum_{j=1}^{n^{QB}} \frac{n_{Re(i),QB(j)}}{n_{Re(i)}} \hat{\beta}_j^{[a]} - \sum_{k=1}^{\kappa} \int \hat{m}_k^{[a+1]}(u) \hat{f}_k^i(u) du.$$

5. Impose the constraint for $i = 1, \ldots, n^{Re}$

$$\hat{\alpha}_i^{[a+1]} = \hat{\hat{\alpha}}_i^{[a+1]} - \sum_{i=1}^{n^{Re}} \frac{n_{Re(i)}}{n} \hat{\hat{\alpha}}_i^{[a+1]}.$$

6. For $j = 1, \ldots, n^{QB}$, calculate

$$\hat{\beta}_j^{[a+1]} = \tilde{\beta}_i - \sum_{i=1}^{n^{Re}} \frac{n_{Re(i),QB(j)}}{n_{QB(j)}} \hat{\alpha}_i^{[a+1]} - \sum_{k=1}^{\kappa} \int \hat{m}_k^{[a+1]}(u) \, \hat{f}_k^j(u) du.$$

7. If a predetermined convergence rule is satisfied, stop. Otherwise set a to $a+1$ and go back to step 2.

We use the convergence criterion proposed in Nielsen and Sperlich (2005),

$$\frac{\sum_{k=1}^{\kappa} \left|\left|\hat{m}_k^{[a+1]}(x_k) - \hat{m}_k^{[a]}(x_k)\right|\right|^2}{\sum_{k=1}^{\kappa} \left|\left|\hat{m}_k^{[a]}(x_k)\right|\right|^2 + 0.0001} < 0.0001,$$

however, any reasonable choice can be used.

A.2 Tables and Figures

Figure 2: Distribution of Player Ability

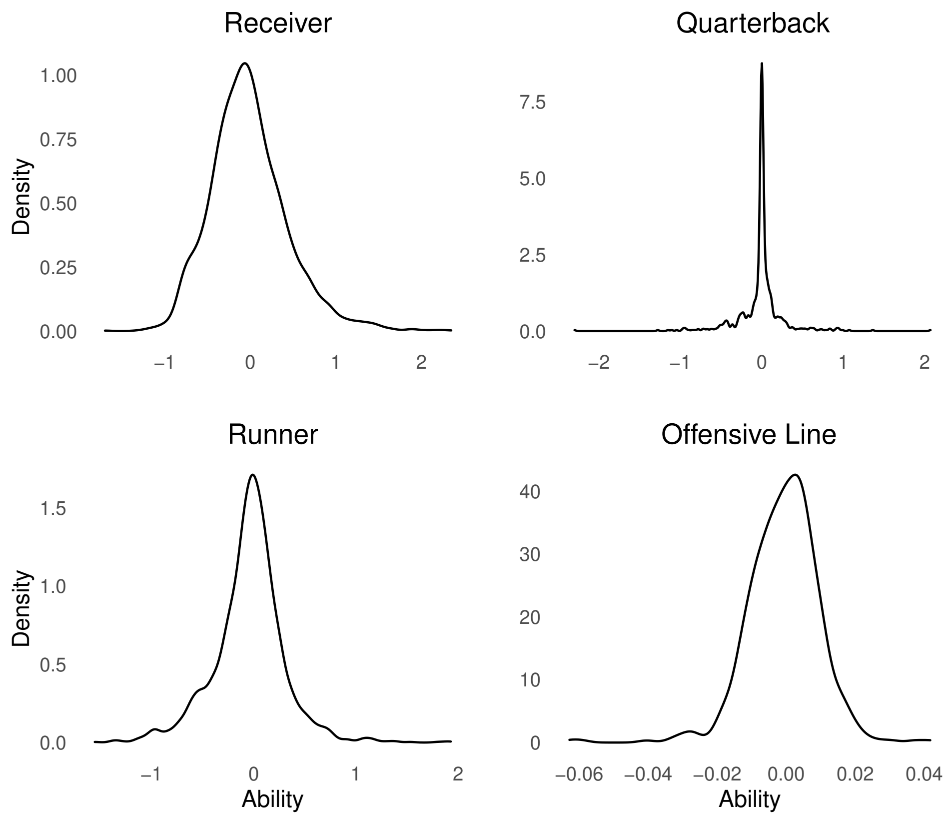


Table 5: Regression Results (Players with at Least Five Attempts)

	Dependent variable:		
	Efficiency	EPA	Win
	(Tobit)	*(OLS)*	*(Probit)*
Maximum Bonus × Play Caller	-0.13^{***}	-0.09^{*}	-0.04^{**}
	(0.02)	(0.05)	(0.03)
Maximum Bonus	0.03^{**}	0.09^{**}	0.10^{***}
	(0.01)	(0.03)	(0.02)
Play Caller	0.04	0.07	−0.06
	(0.05)	(0.11)	(0.06)
Pay	−0.01	-0.07^{**}	0.07^{***}
	(0.02)	(0.04)	(0.02)
Assistant Pay	0.03^{**}	0.06	-0.04^{**}
	(0.02)	(0.04)	(0.02)
Assistant Maximum Bonus	0.01	−0.02	−0.001
	(0.01)	(0.02)	(0.01)
Play Number	0.0001	0.01	-0.02^{***}
	(0.003)	(0.01)	(0.005)
Attendance	0.01^{**}	0.03^{***}	0.09^{***}
	(0.005)	(0.01)	(0.01)
Home	-0.01^{*}	0.07^{***}	0.11^{***}
	(0.01)	(0.02)	(0.01)
Field Position	0.001	0.01	-0.02^{***}
	(0.003)	(0.01)	(0.004)
Points Difference	-0.01^{*}	−0.01	0.24^{***}
	(0.01)	(0.01)	(0.01)
Winning	0.02^{*}	0.06^{**}	0.22^{***}
	(0.01)	(0.03)	(0.02)
Options Variance	-0.32^{***}	0.01	-0.02^{***}
	(0.004)	(0.01)	(0.004)
Options Number	-0.07^{***}	−0.001	0.01^{***}
	(0.003)	(0.01)	(0.005)
Log Sigma	-0.30^{***}		
	(0.01)		
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Regression Results (Maximum Bonus Determinants)

	Dependent variable:
	Maximum Bonus
Play Caller	0.07
	(0.23)
Pay	0.04
	(0.18)
Assistant Pay	0.27^{*}
	(0.16)
Assistant Maximum Bonus	0.13
	(0.08)
Play Number	0.0004
	(0.003)
Attendance	0.01
	(0.01)
Home	-0.01
	(0.01)
Field Position	-0.001
	(0.002)
Points Difference	-0.01^{*}
	(0.01)
Winning	0.02^{**}
	(0.01)
Options Variance	0.004
	(0.004)
Options Number	0.002
	(0.01)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Regression Results (Without Coach Fixed-Effects)

	Dependent variable:		
	Efficiency	EPA	Win
	(Tobit)	*(OLS)*	*(Probit)*
Maximum Bonus × Play Caller	−0.005	−0.002	−0.01
	(0.01)	(0.01)	(0.01)
Maximum Bonus	−0.01*	0.04	0.10***
	(0.004)	(0.01)	(0.005)
Play Caller	−0.01	0.01	−0.01
	(0.01)	(0.02)	(0.01)
Pay	−0.02***	−0.04**	−0.03***
	(0.01)	(0.01)	(0.01)
Assistant Pay	−0.002	−0.001	−0.03***
	(0.01)	(0.01)	(0.01)
Assistant Maximum Bonus	−0.001	−0.0001	−0.02***
	(0.004)	(0.01)	(0.005)
Play Number	−0.0002	0.02**	−0.02***
	(0.003)	(0.01)	(0.004)
Attendance	−0.01***	0.03***	0.06***
	(0.004)	(0.01)	(0.005)
Home	0.01*	0.06***	0.10***
	(0.01)	(0.01)	(0.01)
Field Position	−0.0005	0.001	−0.02***
	(0.003)	(0.01)	(0.004)
Points Difference	0.01	−0.01	0.24***
	(0.01)	(0.01)	(0.01)
Winning	−0.02*	0.06***	0.22***
	(0.01)	(0.02)	(0.01)
Options Variance	0.32***	0.03***	−0.01***
	(0.003)	(0.01)	(0.004)
Options Number	0.07***	−0.01	0.02***
	(0.003)	(0.01)	(0.004)
Log Sigma	−0.19***		
	(0.01)		
Note:			*p<0.1; **p<0.05; ***p<0.01

References

[1] Ai, C. and Norton, E.C. (2003) Interaction terms in logit and probit models. *Economics Letters*, 80(1), pp.123-129.

[2] Ariely, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009) Large stakes and big mistakes. *Review of Economic Studies*, 76(2), pp. 451-469.

[3] Barros, C.P., Peypoch, N. and Tainsky, S. (2014) Cost efficiency of French soccer league teams. *Applied Economics*, 46(8), pp.781-789.

[4] Berri, D.J. (2008) A Simple Measure of Worker Productivity in the National Basketball Association. *The Business of Sport*, 3, pp.1-40

[5] Bradbury J.C. (2008) Statistical Performane Analysis in Sport. *The Business of Sport*, 3, pp.41-56

[6] Bryson, A., Buraimo, B. and Simmons, R. (2011) Do salaries improve worker performance? *Labour Economics*, 18(4), pp.424-433.

[7] Camerer, C., Loewenstein, G. and Prelec, D. (2005) Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1), pp.9-64.

[8] Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2011) Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2), pp. 238-249.

[9] Carter, V. and Machol, R.E. (1971) Operations research on football. *Operations Research*, 19(2), pp.541-544.

[10] Cooper, M., Gulen, H. and Rau, P.R. (2016) Performance for pay? The relation between CEO incentive compensation and future stock price performance. Unpublished Manuscript.

[11] Frey, B.S. and Jegen, R. (2001) Motivation crowding theory. *Journal of Economic Surveys*, 15(5), pp.589-611.

[12] Fryer, R.G. (2013) Teacher incentives and student achievement: Evidence from New York City public schools. *Journal of Labor Economics*, 31(2), pp.373-407.

[13] Hakes, J.K. and Sauer, R.D. (2006) An economic evaluation of the Moneyball hypothesis. *Journal of Economic Perspectives*, 20(3), pp.173-186.

[14] Hastie, T.J. and Tibshirani, R.J. (1990) Generalized additive models. *Statistics and Applied Probability.*

[15] Lavy, V. (2009) Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(5), pp.1979-2011.

[16] Lenten, L.J. (2017) Racial discrimination in umpire voting: an (arguably) unexpected result. *Applied Economics*, 49(37), pp.3751-3757.

[17] MacKinnon, J. G. and White, H. (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3), pp. 305–325.

[18] Mammen, E., Støve, B. and Tjøstheim, D. (2009) Nonparametric additive models for panels of time series. *Econometric Theory*, 25(2), pp.442-481.

[19] McGraw, K.O. and McCullers, J.C. (1979) Evidence of a detrimental effect of extrinsic incentives on breaking a mental set. *Journal of Experimental Social Psychology*, 15(3), pp.285-294.

[20] Mullen, K.J., Frank, R.G. and Rosenthal, M.B. (2010) Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics*, 41(1), pp.64-91.

[21] Savage, D.A. and Torgler, B. (2012) Nerves of steel? Stress, work performance and elite athletes. *Applied Economics*, 44(19), pp.2423-2435.

[22] Simmons, R. and Berri, D.J. (2011) Mixing the princes and the paupers: Pay and performance in the National Basketball Association. *Labour Economics*, 18(3), pp.381-388.

[23] Stiroh, K.J. (2007) Playing for Keeps: Pay and Performance in the NBA. *Economic Inquiry*, 45(1), pp.145-161.

[24] Yu, R. (2015) Choking Under Pressure: The Neuropsychological Mechanisms of Incentive-Induced Performance Decrements. *Frontiers in Behavioral Neuroscience,* 9, pp.19.