Criminals on the Field:
A Study of College Football*

Radek Janhuba† (CERGE-EI)‡ & Kristýna Čechová (IES FSV UK)§

February 28, 2018

Latest version:
http://www.radekjanhuba.com/files/CotF_Janhuba_Cechova.pdf

Abstract

Economists have found mixed evidence on what happens when the number of police increases. On the one hand, more law enforcers means a higher probability of detecting a crime, which is known as the monitoring effect. On the other hand, criminals incorporate the increase into their decision-making process and thus may commit fewer crimes, constituting the deterrence effect. This study analyzes the effects of an increase in the number of on-field college football officials, taking players as potential criminals and officials as law enforcers. Analyzing a novel play by play dataset from two seasons of college football, we report evidence of a monitoring effect being present in the overall dataset. This effect is mainly driven by offensive penalties which are called in the area of jurisdiction of the added official. Decomposition of the effect provides evidence of the presence of the deterrence effect in cases of penalties with severe punishment or those committed by teams with moderate to high ability, suggesting that teams are able to strategically adapt their behavior following the addition of an official.

JEL codes: H43, K14, Z29

Keywords: Football, Official, Crime, Deterrence

*We would like to thank Randall Filer, Jan Hanousek, Stepan Jurajda, participants in a CERGE-EI Brownbag seminar, participants in the ESEA 2017 Paderborn conference, and two anonymous referees for helpful comments and suggestions. All remaining errors are our own. This study was supported with institutional support RVO 67985998 from the Czech Academy of Sciences.

†Corresponding author. Email: radek.janhuba@cerge-ei.cz

‡CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic.

§Institute of Economic Studies, Faculty of Social Sciences, Charles University.

1 Introduction

What is the effect of increasing the number of police on crime rates? Based on the economic model of crime established by Becker (1968), the decision to engage in criminal activities depends on the expected utility of committing a crime. Specifically, potential criminals make their decision based on possible benefits, costs (punishment), probability of conviction, and considering their individual specific characteristics such as education.

An increase in the number of police can increase the probability of being caught and therefore convicted. If this increase is unobserved by potential criminals, it leads to an increase in reported crime rates, constituting a monitoring effect. However, potential criminals will likely observe an increase in the number of police. They will therefore incorporate it into their decision making process, change their behavior, and decrease the number of crimes committed (as their expected utility has decreased). This is the deterrence effect. As the monitoring and deterrence effects have opposite directions, the total effect of increasing the number of police on reported crime rates can be either positive or negative depending on the magnitude of each effect.

Our study looks at sports as an environment in which players are potential criminals and officials take the role of law enforcers.[1] Examining a novel play-by-play dataset, we evaluate the effects of increasing the number of officials from seven

[1]For the purpose of keeping the terminology clear, we abstain from using the term *referee* for a person observing the game and policing the rules. Instead, the term *official* is used. The reason is that in our context there are seven or eight officials on the field, and the one in charge of the whole officiating crew is called the *Referee*. Throughout the study, we identify this official using the term *Referee* (with capital R).

to eight in the 2014 and 2015 National Collegiate Athletic Association (NCAA) football seasons.[2] To the best of our knowledge, this is the first study to examine this policy change on a nation-wide dataset.

In the sports context,[3] the economic model of crime established by Becker (1968) has been examined by several studies modeling fouls committed by players and the number of officials. McCormick & Tollison (1984) show that adding a third official in college basketball led to a decrease in the number of penalties called. Although Hutchinson & Yates (2007) discovered that McCormick & Tollison's results are erroneous due to a coding mistake, the corrected results still present evidence in favor of the existence of the deterrence effect (McCormick & Tollison, 2007).

Levitt (2002) and Heckelman & Yates (2003) analyze an experiment in the National Hockey League where, during the 1999-2000 season, games were observed by either one or two referees.[4] Both papers find that the number of penalties increased and thus show that the monitoring effect was stronger than the deterrence effect (if there was any deterrence effect at all). Levitt (2002) argues that the change in the probability of detection was too small to result in an observable deterrence effect. Heckelman & Yates (2003) concludes that breaking rules in sports might not be well thought out but rather impulsive.

The sports policy evaluations closest to ours were carried out by Kitchens (2014) and Kitchens, Makofske & Wang (2017). Kitchens (2014) analyzes a natural ex-

[2]Note that throughout this study, the word *football* refers specifically to American football.

[3]Models of actual criminal behavior have been examined by several studies, many of which, however, suffer from endogeneity. General studies on crime are not reviewed in this study which is focused on testing the deterrence effect in a sports environment. For a thorough review of the economics of crime, see Paternoster (2010) or Chalfin & McCrary (2014).

[4]In hockey terminology, the term *official* is not widely used. Instead, the game is supervised by referees and linesmen.

periment in the National Football League, which moved the position of the official known as the Umpire from behind the defense to behind the offense, keeping the number of officials fixed at seven.[5] Their results reveal that, after the change in the spatial distribution of officials, the number of penalties called on offense increased by 14 % while the number of penalties called on defense decreased by 17 %.

Interestingly, the eighth official added in college games, the Center Judge, was added to the same spot as the new NFL Umpire position, while the college Umpire stayed in his original spot. Thus, our results may be viewed as complementary to results of Kitchens (2014) in the sense that the policy change we analyze added an official to a specific location, while Kitchens's analysis combined this intervention with removal of an official from a different location.

Kitchens et al. (2017) study the policy change we analyze (see Section 2.2 for a description). On the dataset from the 2012 and 2013 seasons and studying games played by teams that were in the Big 12 Conference, they find evidence of the monitoring effect and limited support for the existence of the deterrence effect. Our study examines the 2014 and 2015 seasons and extends the sample to include all FBS football games.[6]

The contributions of our study are threefold. First, we add to the empirical literature on the strength and existence of monitoring and deterrence effects. By identifying specific types of penalties, we can isolate the two effects. Second, our results indicate that there is a strategic interaction of teams following the policy

[5]The experiment may be viewed as natural due to the fact that the primary reason for moving the Umpire's position was his safety, which is unrelated to the number of penalties called.

[6]FBS is the highest level of college football played in the United States.

change. Third, this is the first study to examine the policy change in question on a nationwide dataset.[7]

Our results indicate the presence of the monitoring effect in the overall dataset. This result is strengthened by performing a decomposition based on the area of officials' jurisdiction. We also find evidence of the existence of the deterrence effect in two scenarios. First, we find an indication of the deterrence effect in cases of penalties carrying severe punishments. This may be explained by teams adapting their behavior as a response to the policy change. Second, we find limited evidence of the deterrence effect present in cases of non-severe penalties when only teams with moderately high (albeit not the highest) ability are considered. This may indicate that only teams at a relatively high playing level are able to strategically change their behavior.

The remainder of this study is structured as follows. Section 2 provides a brief introduction to the rules of football and to the intervention. Section 3 describes the dataset. Section 4 considers the methodology used. Section 5 presents results. Section 6 concludes.

2 Football Specifics and Intervention Details

This section first introduces the sport of (American) football and its specifics that are important for this study. It then describes the intervention and discusses its implications. Readers familiar with the game of football may prefer to skip the next section and proceed directly to Section 2.2.

[7]Thus, our study may be seen as complementary to Kitchens et al. (2017), who examine the same policy change, but only for the Big 12 conference and for the period before our sample starts.

2.1 The Game of Football

Football is a collective sport played with 11 players on two teams on a rectangular field divided by lines into a grid. The last zone on each end of the field is known as the end zone.

The game is conducted in short consecutive plays usually lasting only seconds. After each play, the ball is placed either on the spot where it was at the end of the play, or the spot where the previous play started, depending on the outcome of the play. The team which initiates the ball into play is called the *offense*, and its objective is to get the ball into the opponent's end zone in order to score. The opposing team protects its end zone in order to keep the offense from scoring, and is called the *defense*.

When a team is awarded the ball, it has four opportunities (*downs*) to move the ball at least 10 yards closer to the opponent's end zone. If the offense succeeds, the down count resets and the offense again has a first down and 10 yards to go.[8] If the offense fails to achieve the first down during the four attempts, the ball is turned over to the defense at the spot where the fourth attempt ended.[9] The defense is then awarded a 1st down and hence becomes the offense, and vice versa.

The goal of the game is to score more points than the opposing team. Kicking the ball through the uprights of the "Y" shaped goal results in a field goal worth 3

[8]For example if the situation is labeled a *2nd* (down) & *5*, the team has second down and must advance the ball at least 5 yards to get the first down. If the team advances the ball 3 yards only, the next down will be labeled the *3rd* & *2*. If the team advances the ball 10 yards, they will get *1st* & *10* at the spot where the play ends.

[9]Teams will very rarely attempt to get the first down on a fourth down. Instead, they usually elect to try a field goal (see the next paragraph) or *punt* the ball, in which case they kick it towards the opponent's end zone so that the other team will need to gain a larger distance to score.

points. A touchdown worth 7 points is scored by moving the ball into the opponent's end zone by either carrying it there or catching it there.[10] Last but not least, a safety worth 2 points is awarded to the opposing team if a team is stopped in its own end zone (this occurs very rarely).

The games are governed by seven, or recently eight, officials.[11] These officials observe the game and if they see rule violations,[12] they throw a yellow flag to indicate a penalty. After the play ends, they confer together and then the Referee (the official responsible for the whole crew) informs the teams and spectators of their decision. The usual form of penalty is a loss of 5, 10, or 15 yards, according to the severity of the foul.[13] The penalty is then assessed against the fouling team and the down is repeated.[14]

2.2 Change in the Number of Officials

We analyze a policy change where the number of officials overseeing football games increased from seven to eight. The intervention was implemented gradually over three seasons. In the 2013 season, eight-member officiating crews oversaw exclusively games governed by the Big 12 conference. In the 2014 season, an additional three

[10]Technically, the scoring team receives 6 points for a touchdown. Afterwards, it attempts one more play (so called "extra point" or "try") for which it can receive one point for kicking a field goal, an outcome that happens almost all the time, or two points in case it scores another touchdown. An unsuccessful try for either a field goal or a touchdown means that the team receives 6 points for the touchdown.

[11]Note that this holds only for the highest level of college football games. Lower level college and professional games are governed by seven officials.

[12]Although the basic rules of the game are quite simple, the specifics of play are governed by a complex set of rules (e.g., the 2016 official NCAA football rule book contains 218 pages of text).

[13]In specific circumstances, the penalty can shorten in distance by taking form of half the distance to the goal line or by placing the ball on the spot where the foul was committed.

[14]For example, if the offense commits a holding foul on 2nd & 10 which results in a gain of 15 yards, the gain is canceled and the next down will be 2nd & 20. Note that in some specific cases, the penalty can also include a loss of a down for offensive penalties or an automatic 1st down for defensive infractions.

conferences[15] adopted the same rule change, while in 2015 it was applied to the whole FBS. The gradual introduction enables us to study the intervention as a natural experiment.

The first policy change of this type since 1983 was adopted as a response to an increase in the speed of the game in the previous years, and related issues.[16] The officials began to have difficulty preparing for the next play quickly enough to assure proper observation of the game. Player safety and potential holes in coverage were also widely discussed topics.[17]

Generally, officials have divided areas of coverage, meaning that each official has a specific area to observe and detailed instructions on what types of fouls to watch for in particular. Holes in this coverage meant that specific fouls were missed due to the seven officials not being able to observe all the actions taking place on the field. Note that while no official is strictly restricted from calling fouls that occur outside his area of jurisdiction, the officials are specifically trained to observe their area exclusively, and are actively criticized when they take actions outside of their jurisdiction. Note also that the officials work in crews that remain together for the entire season and they are thus generally well aware of what their colleagues would be doing in each specific situation that may arise during the football game.

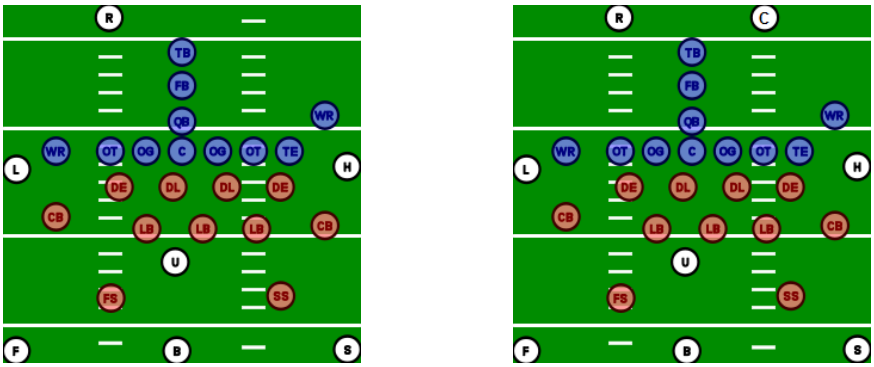
[15]Namely, those were the Atlantic Coast Conference, the Big 10 Conference and the American Athletic Conference.

[16]The increase in pace was associated with the implementation of the 40-second clock. The 40-second clock rule introduced in 2008 sets an interval between the end of a play and the beginning of a new one at no more than 40 seconds. Previously, the clock was only 25 seconds but counting began only after the officials made the ball ready for play. The aim of the rule was to increase the pace of games (Source: http://bleacherreport.com/articles/35981-2008-rule-changes-what-every-fan-needs-to-know).

[17]For examples, see: http://www.cbssports.com/college-football/news/big-12-adds-eighth-official-just-to-keep-up-with-up-tempo-offenses/ or http://www.cbssports.com/college-football/news/sec-to-experiment-with-8-football-officials-but-whats-right-number/

A graphical illustration of the change in the composition of officials is depicted in Figure 1.[18] The added official has been labeled a Center Judge and is positioned in the offensive backfield, behind the offense and to the opposite side from the quarterback than the Referee. Thus, his area of jurisdiction mainly includes fouls in the area of the offensive line (broadly defined) and defensive fouls against the quarterback.

Figure 1: Schematics of the Policy Change



(a) Before (b) After

Source: Wikimedia Commons (see footnote 18 for details)

Even though the main reason for implementing the policy change was unrelated to penalty-related behavior, there is still a potential threat that conferences which voluntarily adopted the policy change in 2014 differed from those that waited until 2015. Therefore, we performed balancing tests for penalty related statistics in the season before the intervention took place.

Specifically, we examined the overall aggregated team levels of penalty related measures from the 2013 season, and analyzed whether their distribution differs for

[18] The scheme of the American football officiating positions has been downloaded from Wikimedia Commons under the Creative Commons Attribution-Share Alike 3.0 Unported license, and was subsequently modified to illustrate the policy change. Author: Derivative work by Zzyzx11 based on the original image by UserB. Detailed information: https://commons.wikimedia.org/wiki/File:American_football_officials_positions.svg

conferences that initiated the eighth official in 2014. We have also checked if the two conference groups differed in the speed of play before the intervention took place. The results of these balancing tests are presented in Table 2.1. Both the t-tests and Kolmogorov-Smirnov tests for equality of distributions suggest that the distribution of balancing characteristics did not differ between conferences which did and did not implement the policy change in 2014. Hence, we conclude that the intervention may be viewed as exogenous to penalty characteristics.

Table 2.1: Balancing Tests

	Control[1]		Treatment[1]		t-test		K-S test	
	Mean	SD	Mean	SD	t-stat	p-val	D stat	p-val
Penalties per game	5.56	1.16	5.69	1.24	-0.52	0.60	0.10	0.96
Penalties per play	0.04	0.01	0.04	0.01	-0.88	0.38	0.13	0.81
Penalty yards per game	48.03	10.73	48.44	11.44	-0.18	0.85	0.10	0.96
Penalty yards per play	0.34	0.07	0.34	0.08	-0.56	0.58	0.11	0.92
Plays per game	143.33	8.98	141.05	7.65	1.30	0.19	0.15	0.67
Plays in season	1817.27	167.03	1797.14	114.48	0.65	0.52	0.20	0.27

Tests performed on aggregate data in the 2013 season. Calculation excludes independent and Big 12 teams.
[1] Treatment includes teams that adopted the intervention in 2014. Control includes teams that did not.
Source: Authors' calculation; Data from http://www.cfbstats.com/2013/team/index.html

Note that in order to keep the decision as clean as possible, our reported results exclude Big 12 and independent teams.[19] Keeping the Big 12 in the dataset would mean that one of the balancing groups would include data that were influenced by the "trial run" of the intervention. Nevertheless, results from balancing tests including the Big 12 teams in the control group are qualitatively the same.[20]

[19] In 2013 there were six so called independent teams. These teams are not governed by any conference.

[20] These results are available upon request.

3 Data

The data on football games have been downloaded from the NCAAsavant.com website.[21] The data include play by play information for NCAA football games in the 2014 and 2015 seasons.[22] Note that the dataset was mainly created as the base of an interactive website and unfortunately does not cover games from the 2013 season (when all games were governed by seven officials), which causes methodological issues (see below).[23]

The dataset includes basic variables about each play, such as which team is on the offense, the type of play, the result of the play, and a detailed text description of the play. This text includes information about penalties called during the play. Therefore, we can identify the penalty type, team, player, and whether the specific penalty was called on the offense or defense.

The aggregated seasonal statistics for each team were obtained from the website of SportSource Analytics.[24] The information from this source contains the aggregates for the number of offensive plays, offensive yards, number of penalties, and penalty yards in 2008 to 2015 seasons. All these variables are available for each team and for its opponents.

The data on officiating assignments were downloaded from the collegiate athletics websites of all 128 universities that were part of the FBS in the 2014 and 2015

[21] http://ncaasavant.com

[22] More precisely, the data present a subsample of football games in each season. In the 2014 season, the missing games seem to be random. In the 2015 season, the dataset covers the first seven weeks of the season.

[23] We tried to contact the owner of the website to get access to codes used to compile the dataset, which would allow us to obtain the same data for the 2013 season (as well as missing games from 2014 and 2015). Our inquiries did not meet with a response.

[24] http://www.cfbstats.com/

seasons. Note that while play-by-play statistics are generally available from sports news websites for our sample period, these servers usually do not include data on which officiating crew supervised the particular game. This information is available in the official game statistics, which the home team is required to collect and upload to the NCAA. The teams then release this official report on their athletics websites. Note that as the officials work in crews that are constant through the entire season, we only record the name of the Referee to identify which officiating group oversaw the specific game.

After matching the three data sources together, the main dataset includes 148,097 plays from 1,011 games. Note that in order to simplify the analysis, we decided to restrict the dataset to basic plays from scrimmage (rushes and passes).[25]

The descriptive statistics for play-by-play data are presented in Table 3.1. The first two rows show the proportions of run and pass plays. The last three rows show the unconditional probability of a penalty occurring, followed by the probability of penalties for offensive holding and roughing the passer. These are the two specific types of fouls we are particularly interested in (see the next section for explanation).[26]

[25]Thus, we eliminate plays involving kicks. Although these are undoubtedly an important part of a football game, the behavior of players during kick plays is substantially different and their inclusion would introduce noise into the analysis.

[26]Note that the number of observations for roughing the passer penalties is approximately half of the number for other variables. This is due to the fact that this type of penalty can only appear in passing plays, while the other types can appear in runs as well as in passes.

Table 3.1: Descriptive Statistics

	Mean	S.D.	Min	Max	N
Running play	0.5107	0.4999	0	1	148,097
Passing play	0.4893	0.4999	0	1	148,097
Any penalty	0.0463	0.2100	0	1	148,097
Offensive holding	0.0125	0.1109	0	1	148,097
Roughing the passer	0.0029	0.0536	0	1	72,462

Source: Authors' calculation

4 Methodology

Utilizing the advantage of having play-by-play information, we examine the probability of a specific penalty being called within every play. The basic model takes the form

$$y_{ighvr} = \lambda_1 \left[\text{eight}_g\right] [2014] + \lambda_2 \left[\text{eight}_g\right] [2015] + \beta X_{ig} + \theta_h + \theta_v + \theta_r + \varepsilon_{ighvr} \quad (1)$$

where the subscript $ighvr$ can be read as "play i in game g of home team h and visiting team v under the supervision of Referee r". The dependent variable y is an indicator equal to one if the specific type of penalty was called within the play. The variables in brackets mark indicators equal to one when the condition described by the inside of the bracket is specified. Specifically, `eight` is an indicator equal to one if the game was supervised by eight officials, and `2014` with `2015` are indicators equal to one if the game was played in a particular season. X is a vector of football specific variables for each play, namely, distance to first down, field position, and indicator variables for down, quarter, and whether the play was a run or pass. Last but not least, θ_h, θ_v, and θ_r are fixed effects for the home team, the visiting team, and the officiating crew represented by the Referee.

The particular regression methodology has been selected in order to perform two

types of comparisons. First, the coefficient λ_1 captures the within-season variation of adding an extra official and thus can capture the immediate adjustment to the new number of officials. Second, the coefficient λ_2 captures the between-season variation and measures the effect of introducing the policy for all games in the 2015 season.

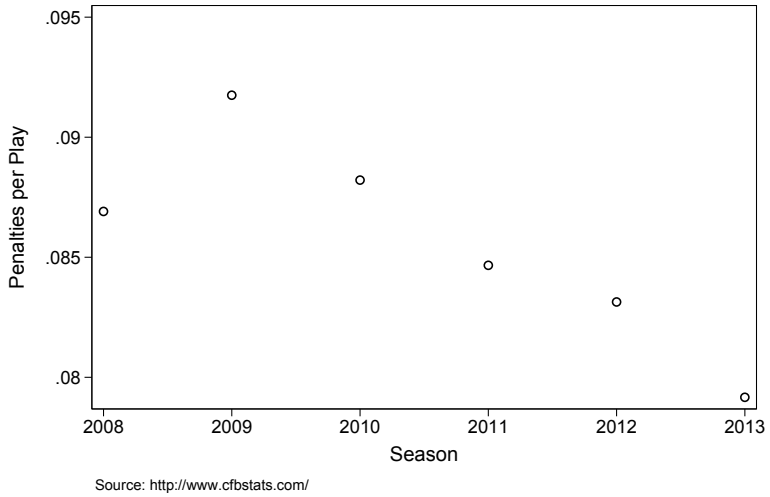
Note that while such methodology would be ideal, it is not plausible to estimate the effect using the standard difference-in-differences framework, as the dataset we possess does not have information on games played in the 2013 season.

Note, moreover, that as the policy change influenced all observations in the second year of the sample, it is difficult to disentangle the effect of the intervention and a potential time trend in the dependent variable. More specifically, if there is a time trend in the dependent variable, one should look at the coefficient and then deduct the time trend from the estimated value of the effect. Figure 2 shows the values of the number of penalties divided by the number of offensive plays and its evolution in years 2008 through 2013.[27] The figure reveals that there is a negative time trend in the number of penalties per play.[28] Therefore, as the regression design inherently assumes that there is a zero time trend, the empirical results will likely tend to underestimate the true effect rather than to overestimate it. In other words, as there is likely a negative time trend in the dependent variable, a potentially positive regression coefficient should arguably be viewed more credibly than it would if it had a negative value.

[27]Note that due to data limitations the unconditional probability of penalties are measured in a different setting and is therefore not comparable to the values in Table 3.1. The reason is that the play by play data for previous seasons is not available and we can therefore only make inferences based on the total number of penalties called on each team including dead-ball fouls such as false starts and/or penalties that are called during kick plays.

[28]The existence of the negative trend is supported by ordinary least squares results of average penalty rates on time.

Figure 2: Pre-Treatment Time Trend in Penalties



In combination with the specific characteristics of several types of football fouls, the intervention allows us to shed light on the difference between the monitoring and deterrence effects. Specifically, these are offensive holdings and roughing the passer penalties, both of which occur predominantly in the area of the new official's jurisdiction. The following paragraphs explain why these two types of penalties can be used for a deeper analysis.

First, holding seems to be the type of foul which is most likely to be influenced by the policy change. Specifically, before the change, the Umpire and the Referee were assigned responsibility for fouls occurring in the area of the offensive line. The basic assignment decomposition was that the Umpire was observing fouls committed by the three interior linemen, while the Referee was observing fouls by both exterior linemen. Clearly, this was often an impossible job, so the Referee was observing only one of the two potential suspects. The introduction of the third observer in the area means that all relevant players can be observed at all times.

Moreover, while it is impossible to prove that there will be no deterrence effect at all, it is also arguably likely that it would be negligible in cases of offensive holding. The reason is that offensive holding occurs in practice when the defensive player outplays his offensive rival, who resorts to illegal holding in order to not allow his opponent to continue to move in the direction of the ball carrier. In fact, coaches often instruct players that, especially in cases of passing plays, they should hold the opponent rather than allow him to continue towards the quarterback, as a holding penalty punishes the team by 10 yards but avoids the potential tragedy of injury to the key player.[29]

Therefore, we find it feasible to suspect that the number of offensive holdings called would have risen following the introduction of the extra official. In terms of the economic model of crime, while there is a higher probability of being caught, the benefits of committing the crime outweigh the potential penalty.

The second type of penalty we are interested is *roughing the passer*, which occurs when the defender hits the quarterback after he has released the ball. The reason is that the second backfield official sees the passer from a second angle and can thus help to cover this safety related foul, therefore the officials are less likely to miss them (constituting a higher probability of detection).

Additionally, such penalties carry an automatic first down for the offense and a risk of the responsible player being disqualified from playing in the remainder of the game in cases of serious misconduct. Therefore, due to severity of the punishment, roughing the passer penalties should arguably be associated with a stronger deter-

[29]The quarterback is the most important player on the team and injury to him may have catastrophic consequences for the team in question.

rence effect. Thus, in terms of the economic model of crime, roughing the passer fouls are crimes with a high punishment.

5 Results

5.1 All Penalties

The results of the linear probability model regressions for all penalties are presented in Table 5.1.[30] The results indicate that although the number of penalties increased

Table 5.1: Linear Probability Model: All Penalties

	(1)	(2)	(3)	(4)
Eight-men crew in 2014	-0.0009	-0.0007	-0.0006	-0.0006
	(0.0018)	(0.0017)	(0.0019)	(0.0021)
Eight-men crew in 2015	0.0026	0.0029*	0.0035**	0.0043**
	(0.0017)	(0.0017)	(0.0017)	(0.0020)
Yards to 1st down		0.0007***	0.0006***	0.0006***
		(0.0002)	(0.0002)	(0.0002)
Field position		-0.0001***	-0.0001***	-0.0001***
		(0.0000)	(0.0000)	(0.0000)
Passing play		0.0276***	0.0273***	0.0273***
		(0.0012)	(0.0012)	(0.0012)
Constant	0.0455***	0.0229***	0.0396***	0.0467***
	(0.0015)	(0.0024)	(0.0131)	(0.0128)
Down and Quarter	No	Yes	Yes	Yes
Teams	No	No	Yes	Yes
Referee	No	No	No	Yes
N	148,097	147,192	147,192	147,192

Standard errors adjusted for 101 clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.

in the 2015 season following the policy change, this increase is not visible in the

[30] We present linear probability model regressions due to the direct interpretation of regression coefficients as marginal effects and their lower computational time required. Robustness of the estimation method is discussed in section 5.5.3.

games supervised by eight officials during the 2014 season, hence the result may not be solely attributable to the presence of the new official. As there has been no other major rule change between the two seasons, one possible interpretation of this result is that the officials may have known that the policy change poses a new issue for the teams and they subsequently "went easy" on the players for the 2014 season. A second possible interpretation is linked with the fact that increasing the number of officials necessarily meant that the newly added official did not have prior experience at the same level.

5.2 Areas of Officiating Coverage

Due to the specific spatial allocation of football officials and their areas of coverage, results on all penalties stacked together may be imprecise as they include fouls which occurred in areas not observed by the new official. Specifically, as the new official was added into the area behind the offense, he would typically be expected to call more penalties on the offense and fewer on the defense.[31] Therefore, we redefined the dependent variable into two separate indicators equal to one when the penalty was called on offense or on defense, and repeated the estimation.

Moreover, in order to analyze the situation in the greatest possible detail, we further devoted our attention to two types of penalties which should arguably be most influenced by the extra official. These are offensive holding and roughing the passer penalties. As explained in Section 4, analysis of these specific penalties should provide insights into the existence of the deterrence effect.

The results are presented in Table 5.2. The first two columns reveal that, as ex-

[31] An exception is roughing the passer penalties which are explored below.

Table 5.2: Linear Probability Model: Area of Coverage

	Offensive Penalties (1)	Defensive Penalties (2)	Offensive Holding (3)	Offensive PI[1] (4)	Roughing the Passer (5)
Eight-men crew in 2014	0.0021	-0.0027	0.0000	-0.0011	0.0014**
	(0.0014)	(0.0016)	(0.0011)	(0.0008)	(0.0007)
Eight-men crew in 2015	0.0045***	-0.0002	0.0020**	-0.0006	0.0006
	(0.0014)	(0.0014)	(0.0009)	(0.0006)	(0.0006)
Yards to 1st down	0.0006***	0.0000	0.0005***	-0.0002***	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Field position	-0.0000	-0.0001***	0.0000	-0.0000*	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	0.0007	0.0267***	-0.0063***		
	(0.0009)	(0.0009)	(0.0007)		
Constant	0.0109	0.0357***	-0.0004	0.0146***	0.0030
	(0.0074)	(0.0086)	(0.0043)	(0.0038)	(0.0049)
N	147,192	147,192	147,192	71,964	71,964

The dependent variable is specified by the column heading.
All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for 101 clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[1] PI stands for "Pass Interference".
Source: Estimation of the model.

pected, the effect can be mainly attributed to increases in offensive penalties. Specifically, while the effect is statistically insignificant in the 2014 season, the probability of a penalty called within a play supervised by seven officials is 0.0215. Hence, the effect of 0.0045 called under the supervision of eight officials in the 2015 season represents an increase of 21.1 percent.

Due to the spatial allocation of the officials, the regression reported in the second column may be viewed as a placebo test. We can see that, as the defensive penalties stay the same following adoption of the eighth official, the placebo test indicates that the increase in penalties is indeed driven by fouls committed on the offense.[32]

[32]Following Kitchens (2014), we also performed a placebo test for defensive holding on runs.

The offensive holding results, in which the deterrence effect is expected to be very small, are presented in the third column. We can see that the results for offensive holding penalties are qualitatively similar to the overall results for all penalties, in the sense that there is an effect of more penalties being called, but only after all of the conferences adopted the eighth official in the 2015 season.[33] Specifically, while the effect is statistically insignificant in the 2014 season, the probability of an offensive holding being called within a play supervised by seven officials is 0.0121, while the effect of 0.002 represents an increase of 16.8 percent called under the supervision of eight officials in the second season. This result is qualitatively similar to the result of Kitchens (2014), who finds an increase of 14% following the relocation of the umpire from behind the defense to behind the offense.

The fourth column provides another placebo test by looking at offensive pass interference penalties, which are arguably the only type of offensive fouls that should not be even theoretically influenced by the eighth official.[34] The fact that the coefficients are insignificant in both periods validates the finding from the regression in the third column.

The results reported in the fifth column seemingly suggest that the number of roughing the passer penalties increased following the policy change, however, only in the 2014 season when some games were still supervised by seven officials. However,

As the coefficient on both variables in question was insignificant at the 5% level, the qualitative implications of this alternative placebo specification stay the same.

[33]We also tried restricting the sample to offensive holdings called during passing plays only. The results are qualitatively identical.

[34]The reason is that when a pass is thrown, all officials except for the Referee and Center Judge look towards where the ball will land. The two remaining officials observe the quarterback, looking out for the roughing the passer penalty. Thus, in the case of offensive pass interference, the presence of the new official carries zero spillover effect.

as discussed in more detail below in section 5.3.2, the coefficient on the variable in the 2014 season is statistically significant only because the 2015 season games are included in the regression and thus influence the regression benchmark. Hence, the positive value of the coefficient in the 2014 season actually picks up the deterrence effect occurring between the two seasons. This may be explained by the fact that in connection with roughing the passer penalties, it is hard to change your behavior when only selected games are supervised by eight officials, while it is more possible to establish a behavioral change between the two seasons.

To sum up, the results in this section present evidence of an overall monitoring effect and suggest the existence of the deterrence effect in the case of crimes with the most serious punishment.

5.3 Experience with the Policy

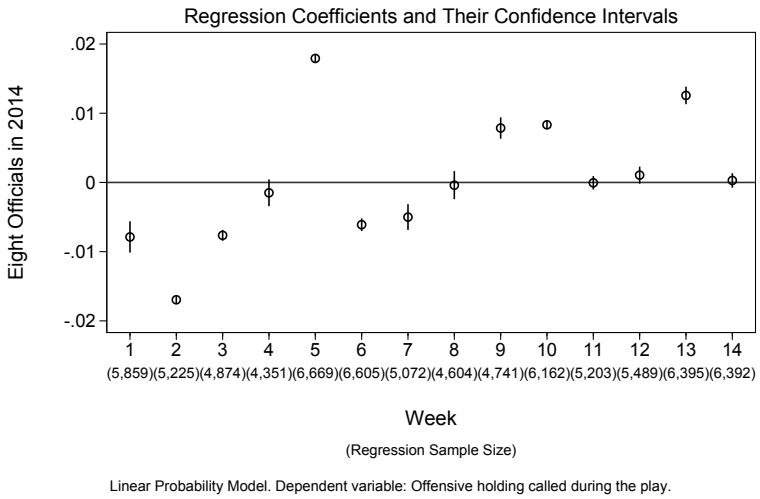
In order to better understand the strategic interactions with regards to the policy change, we now look at regressions estimated on weekly subsamples. Note that due to the policy taking universal effect in 2015, the regressions estimated in this section only look at 2014. Moreover, results in this section exclude games including Big 12 teams from the sample, as these teams had already played under the supervision of eight officials in the 2013 season.[35] Due to the results being extensive in terms of the space needed to show regression outputs, we present the coefficients on the treatment variable graphically. Full results are available upon request.

[35]Including the Big 12 teams does not substantially alter the results.

5.3.1 Offensive Holding

The effects on offensive holding are presented in Figure 3. The figure reveals an interesting pattern during the 2014 season, in which the original regression coefficient was statistically insignificant. The weekly decomposition reveals that, with the exception of one week, there is a clear upward trend in the overall number of reported offensive holdings throughout the 2014 season.

Figure 3: Experience and Offensive Holding



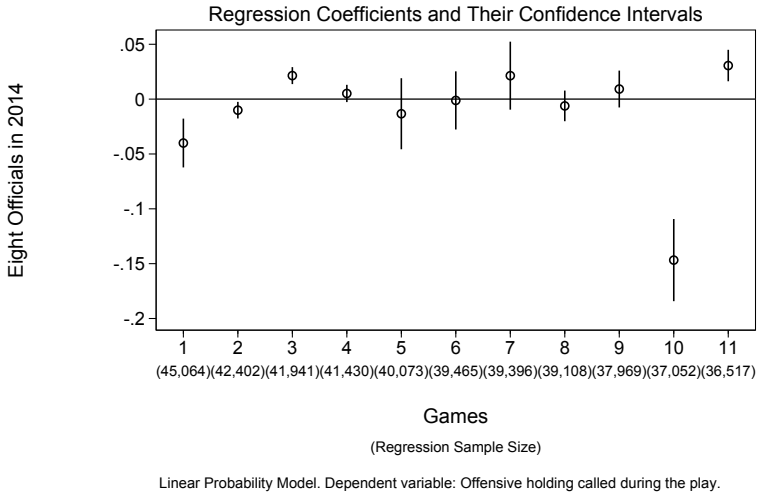
One way to explain this trend is that players are being coached to hold their opponent if they cannot block him legally. With the addition of the extra official in the offensive backfield, players likely started to fear a penalty and thus decreased the number of holding fouls. However, as the season progressed, they were coached not to adjust their behavior and gradually reverted to the overall stable level of fouls committed. Thus, throughout the season, the number of offensive holding fouls increased and the monitoring effect prevailed towards the end of the season.

In order to isolate whether the effect lies in players' ability to adjust their behav-

ior, we also looked at the effects of the policy broken down by the number of games that the crew has officiated up to and including the game in question. For each level of officiating experience, the control group includes all the games supervised by seven officials in the weeks included in the treatment group.

The results of this breakdown are shown in Figure 4. The figure reveals that there is no clear trend in the number of fouls called based on the experience of the officiating crew in question. Therefore, we conclude that the upward trend in offensive holding penalties is likely caused by the strategic interaction of teams in response to the new policy.

Figure 4: Officiating Experience and Offensive Holding

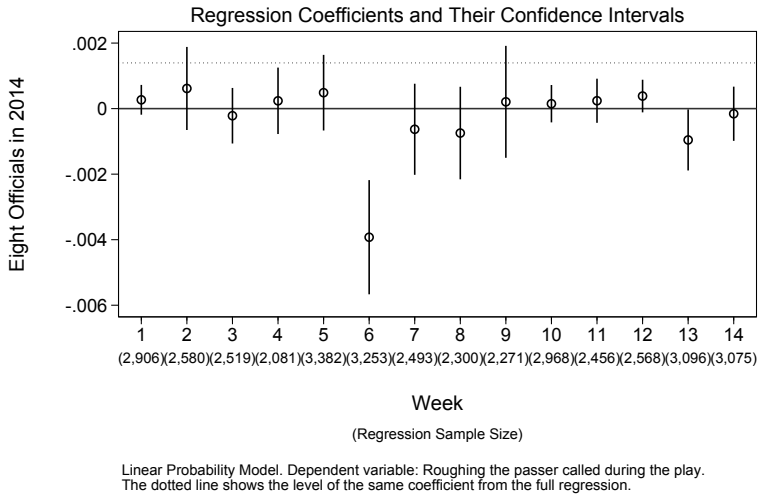


5.3.2 Roughing the Passer

Results on roughing the passer penalties are presented in Figure 5. We can see that there is no clear in-season trend in these cases. More importantly, notice that all of the weekly coefficients are lower in magnitude than the overall effect of the policy in the 2014 season of 0.014. This serves as evidence of the overall coefficient in the

roughing the passer regression reported in column 5 of Table 5.2 actually picking up the deterrence effect of the 2015 season, rather than the monitoring effect of the 2014 season. This is caused by the overall benchmark being influenced by the games from the 2015 season as well.[36]

Figure 5: Experience and Roughing the Passer



5.4 Role of Team Quality

We now extend the analysis to let the effect of the intervention on the two specific types of penalties differ based on team quality. This is motivated by the possibility that high and low skilled teams differ in their game strategies and ability to adjust their behavior in response to the policy change.

In order to distinguish between offensive ability of the teams, we took the total yards gained by each team's offense in the previous season and ranked the teams according to their performance. Analogically, we took the total offensive yards gained by the opposing teams to evaluate defensive abilities. We then defined the

[36]Indeed, when one estimates the roughing the passer regression on the data from 2014 season only, the coefficient on the eight-men crew is statistically insignificant with a p-value of 0.35.

best teams as the top 25 teams in each category. This selection is motivated by the fact that college sports usually rank the best 25 teams overall. The robustness of the number of teams belonging in the top category is discussed in Section 5.5.2.

The results are shown in Table 5.3. The first two columns suggest that the number of offensive holdings already decreased with the addition of the eighth official in 2014, however, only for the teams with a high quality offense. This result indicates the presence of the deterrence effect within teams with high offensive ability.

Table 5.3: Breakdown by Team Quality (LPM)

	Offensive Holding		Roughing the Passer	
	Top 25 Offense (1)	Other Offense (2)	Top 25 Defense (3)	Other Defense (4)
Eight-men crew in 2014	-0.0078***	0.0007	0.0023	0.0016**
	(0.0026)	(0.0011)	(0.0021)	(0.0007)
Eight-men crew in 2015	-0.0033	0.0028***	0.0008	0.0010*
	(0.0025)	(0.0011)	(0.0018)	(0.0006)
Yards to 1st down	0.0007***	0.0005***	0.0001	0.0001
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Field position	0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	-0.0084***	-0.0057***		
	(0.0014)	(0.0007)		
Constant	-0.0445***	0.0035	0.0050	-0.0002
	(0.0106)	(0.0051)	(0.0082)	(0.0049)
N	30,282	116,910	15,220	56,744

Columns are separated by the rankings based on own (opponents') yards gained (allowed) in the previous season.
All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for 84 (Top 25) or 101 clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.

The third and fourth columns report the roughing the passer analysis broken

down to whether the defensive team belongs to the top 25 teams. The effect is insignificant for the teams with the highest defensive quality. This can either mean that there was no effect for these teams, or that (similarly as with offensive holding) the high-quality teams alter their behavior. However, given that the overall roughing the passer rates are the same across the two categories of teams, the difference between the two coefficients could arguably be caused by teams' ability to change their strategic behavior after the policy change. In other words, if there is a deterrence effect, it is likely present for the relatively high skilled teams.

5.5 Robustness Checks

5.5.1 Match Pair Fixed Effects

In this section we perform alternative specification of the regressions. Instead of including a set of fixed effects for the home team and a second set of fixed effects for the visiting team, we keep only those games in which teams played against each other in both years and include a fixed effect for a match-pair combination, ignoring which team played at home and which was on the road.

The results are reported in Table 5.4. Interestingly, all coefficients almost double in magnitude. Moreover, the results for all penalties and offensive holding are more precisely estimated due to the benchmark being more specifically set, utilizing match-pair fixed effects.[37]

In the case of roughing the passer coefficients, its value is larger in magnitude but less precisely estimated.[38] Nevertheless, the reported value is qualitatively consistent

[37]Note that the term "more precisely estimated" is meant in connection with the absolute value of the coefficient. In other words, it does not correspond to a tighter confidence interval, but rather to a result with a higher statistical significance.

[38]Note that because it was the 2014 coefficient that was significant for roughing the passer

with other results.

Table 5.4: Match Pair Fixed Effects

	All penalties		Offensive Holding		Roughing the Passer	
	Team	Match-pair	Team	Match-pair	Team	Match-pair
	(1)	(2)	(3)	(4)	(5)	(6)
Eight-men crew in 2014	-0.0006	0.0036	0.0000	0.0014	0.0014**	0.0021*
	(0.0021)	(0.0039)	(0.0011)	(0.0015)	(0.0007)	(0.0011)
Eight-men crew in 2015	0.0043**	0.0082***	0.0020**	0.0044***	0.0006	0.0014
	(0.0020)	(0.0029)	(0.0009)	(0.0012)	(0.0006)	(0.0009)
Yards to 1st down	0.0006***	0.0002	0.0005***	0.0006***	0.0001	0.0001
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Field position	-0.0001***	-0.0001**	0.0000	0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	0.0273***	0.0283***	-0.0063***	-0.0069***		
	(0.0012)	(0.0019)	(0.0007)	(0.0011)		
Constant	0.0192	0.0288*	-0.0062	0.0072	0.0083**	0.0034
	(0.0154)	(0.0154)	(0.0058)	(0.0063)	(0.0033)	(0.0043)
N	147,192	45,211	147,192	45,211	71,964	22,394

Columns headed by "Team" come from the baseline specification. Columns headed by "Match-pair" only include teams that played each other twice during the sample period.
All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for 81 (Match-pair) or 101 clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.

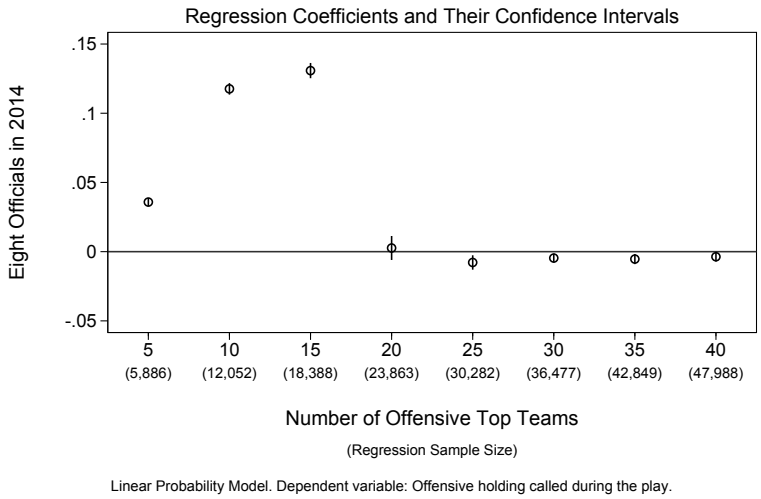
Note also that due to the data being available only for a subsample of games, the reduction of the sample size is substantial. The fact that the results hold even after this decrease in the number of observations further confirms the validity of our results.

regressions, the alternative specification with match-pair fixed effects brings less precision into the estimation of this coefficient. The reason for this is that due to the structure of the competition no two teams played against each other twice in one season.

5.5.2 Number of Top Teams Considered

Even though it is customary to rank the top 25 teams in college sports, the choice of splitting the sample to the best 25 teams remains arbitrary. The sensitivity of the coefficient on eight officials in the second season based on the number of top offensive teams is depicted in Figure 6. Full regression results are presented in Table A.2 in Appendix A.

Figure 6: Offensive Holding and Number of Top Teams

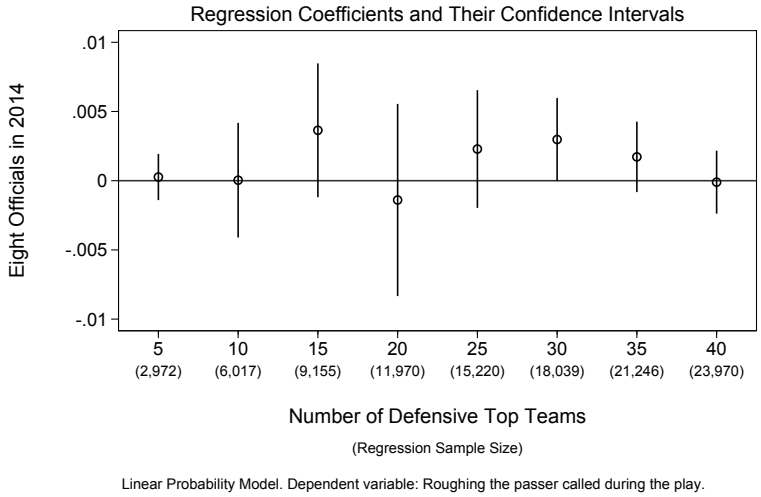


We can see that the result found for the top 25 teams holds if we relax the condition for the 25 teams towards a higher number, but, somewhat surprisingly, does not hold if we constrain the estimation to only the few teams with the strongest offense. Even more surprisingly, analysis of the top 5, 10 and 15 teams suggests exactly the opposite scenario for these teams. A possible explanation for this is that the best teams are so good that they do not need to adjust their behavior in the sense that they are not afraid of a penalty being called. Alternatively, it may be that the top 15 teams are the most skilled and aggressive, hence the addition of the

extra official means that there is a stronger monitoring effect.

The sensitivity of the coefficient for the effect during the 2014 season on the roughing the passer penalties is examined in Figure 7. The full regression results are reported in Table A.3 in Appendix A.

Figure 7: Roughing the Passer and Number of Top Teams



We can see that in the roughing the passer case the interpretation of a possible presence of the deterrence effect holds regardless of how many of the top defensive teams are considered. Interestingly, the very top teams show an increase in the number of roughing the passer penalties in the second season (see Table A.3).

5.5.3 Logit Specification

Due to the direct interpretation of regression coefficients and the less computational time required, all results reported in the main text come from a linear probability model. In order to check for robustness to alternative functional form, the analysis was re-estimated using a logit specification.

Note that, unfortunately, some of the specifications are not estimable with a logit

specification while keeping the full set of fixed effects. The reason is that there are either a very few observations where the dependent variable is equal to one (such as in the case of the roughing the passer penalties) and/or the sample size is not large enough to include the full sets of fixed effects into the analysis.[39]

In order to overcome this issue, we re-estimated these models without the Referee fixed effects. The problem, however, persisted in the specific case of defensive team quality regressions, where there was a combination of small sample size and very small proportion of penalties in the sample. Therefore, in these regressions we kept the Referee fixed effects in the model but excluded team fixed effects. In all cases where the logit estimation was performed using a different set of fixed effects, we re-estimated the same regression using the linear probability model as well.

The comparison of the marginal effects can be found in Appendix B. All results are qualitatively identical to their linear probability model counterparts. Therefore, we conclude that there is no severe functional form specification issue in the analysis.

6 Conclusion

This study evaluates the effects of the policy change of increasing the number of collegiate football officials from seven to eight in the highest level of NCAA football. Comparing our results with the previous literature, this is the first study to find evidence of both the monitoring and deterrence effect on a nationwide dataset.

Analyzing a play by play dataset from two seasons of college football games, we report evidence of a monitoring effect being present in the overall dataset. Moreover,

[39]Including the full set of fixed effects causes the model likelihood to be flat, hence convergence is not achieved when estimating the full form of the model.

analysis of offensive holding and roughing the passer penalties, which constitute misconduct that is especially likely to be observed by the added official, also suggest that there is a monitoring effect present.

We also report evidence of the deterrence effect being present in two scenarios. First, we find an indication of the deterrence effect in the roughing the passer penalties during the second observed season. This is likely caused by between-season changes in team behavior. Second, we find limited evidence of the deterrence effect being present in both types of penalties when only teams with moderate to high ability are considered. This indicates that teams with high (albeit not the highest) skills are able to strategically interact based on the policy change.

The results are robust to alternative specification of fixed effects, functional form of the estimation, and the number of teams considered in the relatively high-skilled group.

References

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The Economic Dimensions of Crime* (pp. 13–68). Springer.

Chalfin, A. & McCrary, J. (2014). Criminal deterrence: A review of the literature. *Journal of Economic Literature*. In press.

Heckelman, J. C. & Yates, A. J. (2003). And a hockey game broke out: Crime and punishment in the nhl. *Economic Inquiry*, *41*(4), 705–712.

Hutchinson, K. & Yates, A. (2007). Crime on the court: A correction. *Journal of Political Economy*, *115*(3), 515–519.

Kitchens, C. (2014). Identifying changes in the spatial distribution of crime: Evidence from a referee experiment in the national football league. *Economic Inquiry*, *52*(1), 259–268.

Kitchens, C., Makofske, M. P., & Wang, L. (2017). Parallel Lives of Hard Hitting Criminals: Evidence from NCAA Football. Working paper, Unpublished.

Levitt, S. D. (2002). Testing the Economic Model of Crime:The National Hockey League's Two-Referee Experiment. *The B.E. Journal of Economic Analysis & Policy*, *1*(1), 1–21.

McCormick, R. E. & Tollison, R. D. (1984). Crime on the Court. *Journal of Political Economy*, *92*(2), 223–35.

McCormick, R. E. & Tollison, R. D. (2007). Crime on the Court, Another Look: Reply to Hutchinson and Yates. *Journal of Political Economy*, *115*(3), 520–521.

Paternoster, R. (2010). How much do we really know about criminal deterrence? *The Journal of Criminal Law and Criminology*, 765–824.

A Appendix A: Number of Top Teams

Table A.1 reports the top teams in the offensive and defensive rankings for each of the two seasons analyzed.

Tables A.2 and A.3 show full regression results from regressions on subsamples on the top teams discussed in Section 5.5.2. Specifically, the coefficients in the first row of Table A.2 are depicted in Figure 6. Figure 7 shows coefficients in the second row of Table A.3.

The sample restriction was performed based on the total number of offensive yards gained in the previous season, with higher numbers of a given team's yards equating to a better offensive ranking. In defensive rankings, the team that allowed the lowest overall number of opponents' offensive yards was ranked the highest.

Table A.1: Overview of Top Teams

Rank	Offensive Top Teams		Defensive Top Teams	
	2014	2015	2014	2015
1	Baylor	Oregon	Louisville	Clemson
2	Oregon	Marshall	Michigan State	Penn State
3	Northern Illinois	Ohio State	Virginia Tech	Stanford
4	Florida State	Baylor	Alabama	Michigan
5	Ohio State	West. Kentucky	Florida	UCF
6	Fresno State	East Carolina	Florida State	Florida
7	Auburn	TCU	Iowa	Louisville
8	Marshall	Boise State	Florida Atlantic	Michigan State
9	Texas A&M	Alabama	Wisconsin	Wisconsin
10	Missouri	Mississippi State	West. Kentucky	LSU
11	Texas Tech	Georgia Tech	Cincinnati	Temple
12	Clemson	Wisconsin	South Florida	App. State
13	Colorado State	Michigan State	TCU	Syracuse
14	Washington	West Virginia	Wake Forest	Arkansas
15	Bowling Green	Arizona	LSU	Boston College
16	BYU	Toledo	Memphis	Virginia
17	Arizona State	Auburn	Bowling Green	Mississippi
18	Georgia	Colorado State	UTSA	Miami (Florida)
19	Wisconsin	Wash. State	North Texas	San Jose State
20	Ball State	Arkansas State	Mississippi State	San Diego State
21	Boise State	Florida State	South Carolina	Florida Intl.
22	Mississippi	Northern Illinois	Oklahoma	Buffalo
23	Cincinnati	UCLA	Tulane	Georgia
24	Indiana	Bowling Green	Penn State	Wake Forest
25	East Carolina	Texas Tech	South Alabama	North Texas
26	Oregon State	Oklahoma	Vanderbilt	UTSA
27	Louisville	BYU	Connecticut	TCU
28	Duke	Cincinnati	Clemson	Akron
29	Arizona	Georgia	Utah State	Houston
30	San Jose State	USC	Kansas State	Virginia Tech
31	Alabama	California	Baylor	Iowa
32	LSU	Texas A&M	Georgia Tech	LA Monroe
33	South Carolina	Nebraska	USC	Texas
34	Oklahoma State	GA Southern	UCF	Connecticut
35	UCLA	Notre Dame	Notre Dame	Memphis
36	Utah State	South Carolina	Texas State	Alabama
37	UCF	Arizona State	Akron	Northwestern
38	Stanford	West. Michigan	Utah	Central Michigan
39	Rice	Fresno State	Pittsburgh	Tulane
40	Wyoming	Pittsburgh	Syracuse	GA Southern

Offensive teams ranked by the most total offensive yards in the specific season.
Defensive teams ranked by the least total opponents' offensive yards in the specific season.
Data from http://www.cfbstats.com/

Table A.2: Offensive Holding Regression: Number of Top Offensive Teams Considered

	Top 5	Top 10	Top 15	Top 20	Top 25	Top 30	Top 35	Top 40
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eight-men crew in 2014	0.0358***	0.1176***	0.1308***	0.0027	-0.0078***	-0.0046***	-0.0053***	-0.0037**
	(0.0015)	(0.0021)	(0.0028)	(0.0044)	(0.0026)	(0.0016)	(0.0018)	(0.0018)
Eight-men crew in 2015	0.0477***	0.1933***	0.2134***	0.0104***	-0.0033	-0.0001	-0.0019	-0.0015
	(0.0015)	(0.0030)	(0.0042)	(0.0034)	(0.0025)	(0.0017)	(0.0017)	(0.0017)
Yards to 1st down	0.0006	0.0008**	0.0010***	0.0008***	0.0007***	0.0007***	0.0006***	0.0007***
	(0.0006)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Field position	0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	-0.0081**	-0.0086***	-0.0075***	-0.0081***	-0.0084***	-0.0091***	-0.0082***	-0.0077***
	(0.0033)	(0.0022)	(0.0018)	(0.0016)	(0.0014)	(0.0013)	(0.0012)	(0.0012)
Constant	-0.0296***	-0.2217***	0.1170***	-0.0244***	-0.0445***	-0.0325***	-0.0306***	-0.0281***
	(0.0057)	(0.0040)	(0.0061)	(0.0077)	(0.0106)	(0.0106)	(0.0107)	(0.0089)
N	5,886	12,052	18,388	23,863	30,282	36,477	42,849	47,988

All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for 48 (Top 5) to 92 (Top 40) clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.

Table A.3: Roughing the Passer Regression: Number of Top Defensive Teams Considered

	Top 5	Top 10	Top 15	Top 20	Top 25	Top 30	Top 35	Top 40
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eight-men crew in 2014	0.0003	0.0000	0.0036	-0.0014	0.0023	0.0030*	0.0017	-0.0001
	(0.0008)	(0.0021)	(0.0024)	(0.0035)	(0.0021)	(0.0015)	(0.0013)	(0.0011)
Eight-men crew in 2015	0.0007	0.0129***	0.0183***	-0.0011	0.0008	0.0007	-0.0001	-0.0014
	(0.0009)	(0.0032)	(0.0038)	(0.0026)	(0.0018)	(0.0016)	(0.0013)	(0.0011)
Yards to 1st down	0.0003	0.0003	0.0003	0.0001	0.0001	0.0000	0.0001	0.0000
	(0.0005)	(0.0003)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Field position	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	-0.0049	0.0016	-0.0144**	0.0153	0.0050	0.0090	-0.0002	0.0035
	(0.0062)	(0.0044)	(0.0059)	(0.0106)	(0.0082)	(0.0071)	(0.0071)	(0.0066)
N	2,972	6,017	9,155	11,970	15,220	18,039	21,246	23,970

All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for 48 (Top 5) to 92 (Top 40) clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.

B Appendix B: Logit Results

Tables in this section present regression results based on alternative functional form using logit estimation. Otherwise, specification of all tables is identical. With the exception of Table B.5, the particular sub-number of all tables corresponds to the sub-numbers of tables in Section 5.[40]

Note that in columns marked by subscripts ^{R} or ^{T} in the heading of the columns, the regression does not include Referee or team fixed effects and is therefore not directly comparable to the appropriate regression in the main text. The reason is that some of the specifications were not estimable using a logit while keeping a the fixed effects due to either a small number of observations with the dependent variable equal to one or insufficient sample size causing the likelihood function to become flat.

As discussed in Section 5.5.3, in order to establish the validity of comparisons in the case described in the previous paragraph, we decided to re-estimate these models without Referee or team fixed effects using both the linear probability model and logit specifications. Results of these regressions are shown in Table B.5.

[40]Thus, for example, Table B.2 corresponds to Table 5.2.

Table B.1: Marginal Effects from Logit Model: All Penalties

	(1)	(2)	(3)	(4)
Eight-men crew in 2014	-0.0009	-0.0008	-0.0007	-0.0002
	(0.0018)	(0.0017)	(0.0020)	(0.0021)
Eight-men crew in 2015	0.0025	0.0028*	0.0035**	0.0046**
	(0.0017)	(0.0017)	(0.0017)	(0.0020)
Yards to 1st down		0.0006***	0.0005***	0.0005***
		(0.0001)	(0.0001)	(0.0001)
Field position		-0.0001***	-0.0001***	-0.0001***
		(0.0000)	(0.0000)	(0.0000)
Passing play		0.0285***	0.0281***	0.0282***
		(0.0013)	(0.0013)	(0.0013)
N	148,097	147,192	147,192	147,192

Standard errors adjusted for 101 clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.

Table B.2: Marginal Effects from Logit Model: Area of Coverage

	Offensive Penalties (1)	Defensive Penalties (2)	Offensive Holding (3)	Offensive PI[1R] (4)	Roughing the Passer[R] (5)
Eight-men crew in 2014	0.0024*	-0.0033*	0.0002	-0.0014	0.0009
	(0.0014)	(0.0017)	(0.0012)	(0.0012)	(0.0013)
Eight-men crew in 2015	0.0047***	-0.0007	0.0022**	-0.0009	-0.0001
	(0.0013)	(0.0015)	(0.0009)	(0.0009)	(0.0012)
Yards to 1st down	0.0005***	-0.0000	0.0005***	-0.0003***	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Field position	-0.0000	-0.0001***	0.0000	-0.0000**	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	0.0008	0.0301***	-0.0064***		
	(0.0009)	(0.0011)	(0.0007)		
N	147,192	146,639	145,402	44,484	43,918

The dependent variable is specified by the column heading.
All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for 101 clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[1] PI stands for "Pass Interference".
[R] Robust regression estimated without Referee fixed effects.
Source: Estimation of the model.

Table B.3: Breakdown by Team Quality (Logit MEs)

	Offensive Holding		Roughing the Passer	
	Top 25 Offense^{R}	Other Offense	Top 25 Defense^{T}	Other Defense^{T}
	(1)	(2)	(3)	(4)
Eight-men crew in 2014	-0.0089***	0.0012	0.0017	0.0001
	(0.0028)	(0.0013)	(0.0040)	(0.0008)
Eight-men crew in 2015	-0.0021	0.0032***	-0.0024	0.0004
	(0.0027)	(0.0011)	(0.0029)	(0.0008)
Yards to 1st down	0.0006***	0.0004***	0.0001	0.0001
	(0.0002)	(0.0001)	(0.0003)	(0.0001)
Field position	0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	-0.0100***	-0.0058***		
	(0.0017)	(0.0007)		
N	25,630	115,353	7,479	45,269

Columns are separated by the rankings based on own (opponents') yards gained (allowed) in the previous season.
All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
^{R} Robust regression estimated without Referee fixed effects.
^{T} Robust regression estimated without team fixed effects.
Source: Estimation of the model.

Table B.4: Match Pair Fixed Effects (Logit MEs)

	All penalties		Offensive Holding		Roughing the Passer	
	Team	Match-pair	Team	Match-pair	Team^{R}	Match-pair^{R}
	(1)	(2)	(3)	(4)	(5)	(6)
Eight-men crew in 2014	-0.0002	0.0024	0.0002	0.0011	0.0009	-0.0000
	(0.0021)	(0.0041)	(0.0012)	(0.0017)	(0.0013)	(0.0027)
Eight-men crew in 2015	0.0046**	0.0079**	0.0022**	0.0046***	-0.0001	-0.0002
	(0.0020)	(0.0031)	(0.0009)	(0.0015)	(0.0012)	(0.0020)
Yards to 1st down	0.0005***	0.0001	0.0005***	0.0005***	0.0001	0.0003
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0003)
Field position	-0.0001***	-0.0001**	0.0000	0.0000	-0.0000	-0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play	0.0282***	0.0294***	-0.0064***	-0.0071***		
	(0.0013)	(0.0021)	(0.0007)	(0.0012)		
N	147,192	45,410	145,402	42,603	43,918	8,003

Columns headed by "Team" come from baseline specification. Columns headed by "Match-pair" only include teams that played each other twice during the sample period.
All columns include the full set of fixed effects for down, quarter, teams, and Referee.
Standard errors adjusted for clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
^{R} Robust regression estimated without Referee fixed effects.
Source: Estimation of the model.

Table B.5: Regressions with Different Sets of Fixed Effects: Comparison of LPM and Logit Marginal Effects

	(1) to (8): Without Referee Fixed Effects								(9) to (12): Without Team Fixed Effects			
	Offensive PI[1]		Rough. the Passer		Rough. the Passer (Match-pair FE)		Offensive Holding (Top 25 Offense)		Rough. the Passer (Top 25 Defense)		Rough. the Passer (Other Defense)	
	LPM (1)	Logit (2)	LPM (3)	Logit (4)	LPM (5)	Logit (6)	LPM (7)	Logit (8)	LPM (9)	Logit (10)	LPM (11)	Logit (12)
Eight-men crew in 2014	-0.0009	-0.0014	0.0007	0.0009	-0.0001	-0.0000	-0.0007	-0.0089***	0.0008	0.0017	0.0001	0.0001
	(0.0008)	(0.0012)	(0.0007)	(0.0013)	(0.0010)	(0.0027)	(0.0023)	(0.0028)	(0.0017)	(0.0040)	(0.0006)	(0.0008)
Eight-men crew in 2015	-0.0007	-0.0009	-0.0000	-0.0001	-0.0001	-0.0002	0.0011	-0.0021	-0.0011	-0.0024	0.0003	0.0004
	(0.0005)	(0.0009)	(0.0005)	(0.0012)	(0.0008)	(0.0020)	(0.0021)	(0.0027)	(0.0011)	(0.0029)	(0.0006)	(0.0008)
Yards to 1st down	-0.0002***	-0.0003***	0.0001	0.0001	0.0001	0.0003	0.0008***	0.0006***	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.0001)	(0.0003)	(0.0001)	(0.0001)
Field position	-0.0000*	-0.0000**	-0.0000	-0.0000	-0.0000	-0.0001	0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passing play							-0.0081***	-0.0100***				
							(0.0014)	(0.0017)				
N	71,964	44,484	71,964	43,918	22,511	8,003	30,282	25,630	15,220	7,479	56,744	45,269

Standard errors adjusted for clusters by the Referee in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Estimation of the model.