AND A HOCKEY GAME BROKE OUT: CRIME AND PUNISHMENT IN THE NHL

JAC C. HECKELMAN and ANDREW J. YATES

We apply the economic theory of crime to the National Hockey League. We analyze a natural experiment in which games during the 1999–2000 season had either one or two referees. We determine the effect of the number of referees on both the number of penalties called and the number of rules infractions committed by players. The results indicate that increasing the number of referees leads to greater enforcement of the rules but does not significantly deter players from committing infractions. (JEL D0, K4)

I. INTRODUCTION

The economic theory of crime predicts that an increase in policing resources will lead to a decrease in the crime rate. Empirically determining the magnitude of this effect has proved difficult, however, because of an endogeneity problem. Although variations in the allocation of policing resources are expected to affect crime rates, the reverse may also hold true (Cornwell and Trumbull, 1994; Levitt, 1997). The endogeneity problem can be avoided to a large degree by studying situations in which changes in the allocation of policing resources occur independent from crime rates.

In an innovative and influential study, McCormick and Tollison (1984) apply the economic theory of crime to rules infractions in a sports contest. They analyze a policy change in the Atlantic Coast Conference (ACC) basketball tournaments. In 1979, the ACC increased the number of referees in the tournament games. McCormick and Tollison analyze the effect of increasing the number of referees on the number of fouls called. Their approach is unlikely to suffer from a severe endogeneity bias because the policy change occurred only once and furthermore occurred between tournament seasons. On the other hand, insofar as rules infractions in sports contests are analogous to criminal activity, McCormick and Tollison measure arrests, rather than crimes.

The effect of an increase in policing resources on arrests cannot be resolved by theory. For a given crime rate, increases in police budgets and forces enable greater monitoring of criminal activity and consequently lead to more arrests. The crime rate, however, is itself a function of the level of policing resources. Rational criminals realize that greater monitoring increases the probability that their actions will result in arrest and may be deterred from committing crimes. This decreases the crime rate. The net effect on the total number of arrests is ambiguous because it depends on whether the monitoring or deterrent effect dominates. McCormick and Tollison find that increasing the number of referees leads to a reduction in the number of fouls called. This suggests that the deterrent effect dominates the monitoring effect. Shortly after their study was published, it was recognized in surveys of both sports economics, by Cairns et al. (1985), and crime, by Cameron (1988).

We examine a natural experiment in sports to further understand the monitoring and deterrent effects. In the 1999–2000 season, the National Hockey League (NHL) games...
had either one or two referees. We find that games with two referees have more penalties called, suggesting that the monitoring effect dominates the deterrent effect. We then use an instrumental variables technique to determine the effect of the number of referees on the number of infractions actually committed by players in the game. We find that the number of referees does not significantly affect the number of infractions committed. This is direct evidence that the deterrent effect is inconsequential in this context. Our results are unlikely to suffer from an endogeneity bias because the variation in the number of referees is independent of the rate of infractions in the individual games.

II. THE NHL EXPERIMENT

Enforcement of the rules in an NHL game is done by referees and linesmen. The NHL has historically utilized one referee and two linesmen. The linesmen are responsible for identifying infractions that result in a stoppage of play, such as icing (slanding the puck from one end of the rink to the other) and off-sides (entering the offensive zone before the puck). The referee is responsible for identifying more severe infractions, such as slashing (swinging a stick at an opponent), hooking (using a stick to impede the progress of an opponent), and fighting (fisticuffs). When one of these infractions is identified by the referee, a penalty is called and the offending player is removed from the ice for a period of time depending on the severity of the penalty. Minor penalties last for two minutes. During this time, the offending player’s team is shorthanded for the duration of the penalty or until the opposition scores, whichever comes first. (The opportunity afforded to the opposition by having relatively more players on the ice is called a power play.) Double minor penalties are similar to minor penalties except that the power play lasts for four minutes. Major penalties last for five minutes, and the opposition may score multiple times during the resulting power play. (Most major penalties are due to fighting.) The most severe penalty is a misconduct. This penalty is usually assessed on top of another penalty. The player is removed from the game for 10 additional minutes (or ejected in the case of a game misconduct), but the opposition does not get additional power play time.

During the 1998–99 season, the NHL experimented by adding an additional referee for 240 games (20% of the season.) For the 1999–2000 season, the number of two-referee games were expanded considerably. In fact, 50 two-referee games were distributed over the 82-game schedule for each of the 28 teams. In total, there were 700 games with two referees and 448 games with one referee. The use of two referees in the 1998–99 season suggests any learning curve should be complete by the start of the 1999–2000 season and enable a deterrent effect, if it exists, to be clearly understood by the players and coaches. Beginning with the 2000–2001 season, all games have two referees. Therefore, our study is exclusive to the 1999–2000 season.

III. CRIMINALS ON ICE

The economic theory of crime posits that criminals consider costs and benefits before deciding whether to commit a crime. The theory applies to the case in which a single party commits the crimes. In a sports contest, there are two parties in an adversarial relationship, and so the theory must be modified to account for this complication. We first describe how the economic theory of crime applies to a single team in a hockey game and then modify this description to include the effects of the behavior of the other team.

At the outset, it is important to distinguish between infractions and penalties. An infraction occurs when a player violates the rules, 2. A separate category of penalties is called penalty shots. A penalty shot occurs when a player is caught committing an infraction against another player who otherwise would have had a breakaway. The offending player is not removed from the ice; rather, the player who was interfered with skates by himself to the goal for a single shot against the goalie. Penalty shots are rare; only 40 penalty shots were awarded during the 1999–2000 season. They are not included in our study.

1. The NHL experiment has attracted the independent attention of several other scholars. Both Allen (2002) and Depken and Wilson (2002) analyze the effect of the number of referees on penalties called but do not make a determination of the effect on the number of infractions committed. Levitt (2002) does consider both penalties called and infractions committed but uses a different technique to estimate infractions than the one employed here.

3. This hypothesis is supported by our data. Neither a position in the season (game number) variable nor an interaction between the position in the season and number of referees variables was significant in regressions explaining penalties called, controlling for the other variables to be detailed.
whereas a penalty occurs when the referee thinks that a player has violated the rules. Infractions and penalties do not always coincide because the referee may make errors. There are two types of errors. Given that an infraction actually occurred, a referee may fail to call a penalty. Given that an infraction did not occur, a referee may call a penalty anyway. Using standard terminology, we call the first error a false negative and the second error a false positive. True positives and true negatives correspond to the cases in which penalties and infractions (or the lack thereof) do coincide. The rate at which infractions are committed is unobservable; we can only measure the rate at which penalties are called. This is determined by the underlying rate of infractions as well as the true-positive and false-positive rates.

Assume for the moment that the infraction rate is kept constant and the number of referees increases. This may lead to either an increase in detection or an increase in accuracy. If detection increases (both the true-positive and false-positive rates increase) then more penalties are called. On the other hand, if accuracy increases (true positives increase but false positives decrease) then the effect on penalties is ambiguous. We assume that false positives are small in magnitude relative to false negatives, and thus we do not need to be concerned with the distinction between detection and accuracy. Under this assumption, and holding the infraction rate constant, an increase in the number of referees leads to an increase in penalties (the monitoring effect).

We use the distinction between infractions and penalties to delineate costs and benefits for a single team in a hockey game. Benefits are a direct function of infractions. In hockey, like many other sports, one benefit of committing an infraction is that it reduces the effectiveness of (or eliminates altogether) a scoring chance for the opponent. For example, a defenseman may use his stick to trip a player to prevent a shot on goal. Hockey is fairly unique, however, in that some infractions do not directly relate to scoring opportunities (either for the opposing or their own team). Such penalties frequently occur when a player known for aggressive play (lovingly referred to by the fans as a “goon”) feels the need (or is instructed by the coach) to protect the star players from being abused. Costs are slightly more complicated. Penalties are a function of the number of infractions as well as the number of referees. Costs are a function of penalties. The cost of a penalty is that the opponent team enjoys a power play for a period of time and is much more likely to score a goal. As mentioned in the previous section, there are differing costs for minor and major penalties. We expect that the monitoring and deterrent effects differ across types of penalties. Although a major penalty carries a larger cost at the margin, it is typically (but not always) offset by a simultaneous major penalty on the opposition, and thus no power play ensues. (Recall that major penalties are due primarily to fighting.) Unlike minor penalties, which are part of the flow of the game and may be missed by the referee(s), a fight is hard to miss. Because players do not try to elude referee detection during a fight, a direct deterrent effect from the presence of an additional referee is unlikely to hold. Given that fights are easily observable, we expect the rate of false positives as well as false negatives for major penalties to be trivial regardless of the number of referees.

Now consider the effects of the other team on the cost and benefit calculation for the current team. The number of infractions committed by the other team primarily affects the benefits of committing an infraction, not the cost of being assessed a penalty. In particular, if the number of infractions (either minor or majors) committed by the other team increases, then it is likely that the benefits of committing an infraction increases. (If the opponent is playing rough, then there is an incentive to play rough as well.) An individual team’s strategy about committing infractions is therefore influenced by expectations about the behavior of the other team. Rather than simply using costs and benefits to determine the optimal strategy for the current team, one must simultaneously determine an equilibrium pair of strategies for both teams. In this equilibrium, the comparative statics of a change in the number of referees includes a direct deterrent effect as well as an indirect “reaction effect.” The deterrent effect was identified before—an increase in the number of referees increases the expected cost of an infraction and hence leads to a decrease in infractions. The reaction effect accounts for the behavior of the other team. Because of the deterrent effect, the other team decreases minor infractions in the presence of an additional referee, and this decreases the benefits of any infraction (minor or major) from the point of view of the current team. Hence, the current team decreases infractions.
In summary, when the number of referees is increased, there are differential effects on major and minor infractions. For minor infractions, there is both a deterrent effect and a reaction effect, and the reaction effect reinforces the deterrent effect. For major infractions, there is only a reaction effect to the other team committing fewer minor infractions (due to the other team’s own deterrent effect). In either case, for minor or majors, the finding of a reduction in infractions from an increase in the number of referees suggests a deterrence effect exists, either for the current team or its reaction to a deterrence effect on the other team. For this reason, when we identify the effect of the number of referees on infractions, we simply refer to this effect as the deterrent effect.4

IV. DATA

Our data are taken from the box scores for the 1999–2000 NHL season. The box scores are available online at www.scoresandstats.com.5 Our unit of observation is a game played by a home team. Our cost and benefit variables include such statistics as effectiveness in defending power plays. To avoid using statistics from a completed game as explanatory variables in a regression explaining the occurrence of penalties in that game itself, we use a moving average of statistics in the previous 15 games for our benefit and cost variables.6 Thus, we lose at least 15 games’ worth of observations for each team at the beginning of the season.7 In addition, from March 22, 2000, to the end of the season, all games had two referees.8 Because the allocation of referees was not random for these games, they may have been viewed differently by the commissioner’s office and also by the players. This suggests that marginal impacts may differ from games earlier in the season, so we drop these games from our sample. We are left with 770 observations.

We measure hockey penalties four ways. First, we follow McCormick and Tollison’s approach by calculating the total number of penalties (TOTAL) each team is called for in a game. We also use the standard NHL statistic of penalties in minutes (PIM), which puts more weight on the more severe penalties. Finally, we also consider the number of minor (MINOR) and major (MAJOR) penalties separately. The correlations among these four measures are shown in Table 1. The total number of penalties is dominated by the more frequent occurrence of minor penalties, whereas PIMs is weighted toward majors (and the rare misconducts). Minor penalties account for 87% of the total number of penalties called and 69% of the penalty minutes assessed.

Table 2 presents the mean value for each penalty measure broken down by the number of referees.9 There were more penalties called in

---

**TABLE 1**

<table>
<thead>
<tr>
<th></th>
<th>PIM</th>
<th>TOTAL</th>
<th>MINOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MINOR</td>
<td>0.66</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>MAJOR</td>
<td>0.73</td>
<td>0.57</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**TABLE 2**

<table>
<thead>
<tr>
<th>Penalty Measure</th>
<th>One Referee (n = 330)</th>
<th>Two Referees (n = 440)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIM**</td>
<td>13.20 (8.39)</td>
<td>14.66 (10.03)</td>
</tr>
<tr>
<td>TOTAL**</td>
<td>5.23 (2.37)</td>
<td>5.73 (2.68)</td>
</tr>
<tr>
<td>MINOR**</td>
<td>4.62 (2.00)</td>
<td>5.01 (2.08)</td>
</tr>
<tr>
<td>MAJOR</td>
<td>0.48 (0.77)</td>
<td>0.55 (0.81)</td>
</tr>
</tbody>
</table>

**Difference in mean significant at 5%.

---

4. The magnitude of the reaction effects may differ across major and minor infractions, and thus we cannot predict the relative size of the overall deterrent effect.

5. This collection is missing box scores, however, for games played on January 10, February 26–28, March 2, and March 5.

6. The 15-game window was chosen arbitrarily but reflects the intuition that performance in recent games is the relevant measure of expected performance in the current game. To check the robustness of our results, we also considered 10- and 20-game moving averages. The statistical significance of our referee variable was not affected, although the level of statistical significance (but not coefficient signs) of some of the control variables were slightly affected.

7. We only include games for which we are able to compute the 15-game moving average for both teams. Because the scheduling does not have every team having played an identical number of previous games, some teams lose a few additional games from our sample.

8. This is approximately game number 72–74 for most teams.

9. Mean values for each penalty measure across all games are presented at the bottom of Table 3.
two-referee games, defined either by the total number of penalties or the number of minors or majors, although the difference in the number of majors is not statistically significant.

There appears to be a significant monitoring effect created by the addition of a second referee. This finding does not, however, shed direct light on the deterrent effect. It may be the case that the deterrent effect exists but is dominated by a much stronger monitoring effect, or perhaps the deterrent effect is altogether absent. We now turn to directly measuring the deterrent effect. In other words, do players actually modify their behavior in games which have two referees?

V. TESTING FOR THE DETERRENT EFFECT

Our empirical representation for applying the economic theory of crime to hockey follows

\[ y^* = a + x_1 b_1 + XB + \mu \]

where \( y^* \) is the number of infractions committed in a game, \( x_1 \) is the number of officials, and \( X \) is a vector of control variables. Although we cannot directly observe the total number of infractions actually committed, we can infer the latent relationship

\[ y = y^* + \varepsilon \]

where \( y \) denotes the penalties called by the referee(s) and \( \varepsilon \) is the error term. As discussed, penalties and infractions do not perfectly coincide. Under our assumption that false positives are small in magnitude relative to false negatives, the error term essentially captures the difference between penalties and infractions due to false negatives.

Because only the penalties called are observable, the standard regression utilized in the sports-crime literature, following from McCormick and Tollison (1984), is of the form

\[ y = a + x_1 b_1 + XB + \mu. \]

An ordinary least squares (OLS) estimate of \( b_1 \) measures the marginal impact of additional referees on penalties called, which represents the combined monitoring and deterrent effect. As it turns out, for the set of controls described below, OLS regressions representing equation (3) merely reinforce the results from the simple difference in means presented in Table 2. We do not present these results, however, because we want to focus attention on isolating the deterrent effect.

Toward that end, suppose for the moment that \( \varepsilon \) was orthogonal to \( x_1 \). Then the OLS estimate of \( b_1 \) would also be an unbiased estimate of the impact of the number of referees on infractions committed. We expect, however, that \( x_1 \) is correlated with \( \varepsilon \) because monitoring errors are a function of the number of referees. To avoid this problem, we use an instrumental variable (IV) technique to estimate the impact of the number of referees on infractions committed. The manner of the NHL experiment suggests a natural set of instruments that would be correlated with the number of referees assigned to a particular game but not the number of penalties called in the game. We describe these instruments after discussing the rest of the data.

The primary variable of interest is the number of referees, which is an indicator variable that takes on the value of one or two. The benefits of committing an infraction are accounted for by two variables. The first variable is the opposing team’s offensive prowess, as measured by its shooting percentage over the 15 previous games. The better the opposing team has been shooting the puck, the greater the benefit of committing an infraction to obstruct or prevent a shot. The second variable is the own team’s defensive prowess as measured by its save percentage over the 15 previous games. The better their own save percentage, the less need to commit infractions. Likewise, the cost of infractions are accounted for by two 15-game running average variables, based on power play proficiency. The cost of an infraction is that it gives the opponent team a power play. The better the conversion rate of the opposing team’s recent power play chances, the greater the potential cost in being called for a penalty. The better the team has been killing opposition power play chances, the less potential cost.

We also include characteristics of the players and coaches. Using box score information, we identify which players from a team’s active roster actually played in the game. Combining this with roster information from the NHL Official Guide and Record Book yields the average height, weight, and age for players
in each game for each team. The differential in these variables between the two teams is included in the regressions as well as the differential in coaching experience (in years). As mentioned earlier, a unique feature of hockey is the presence of goons, who are known for their physical play. We identify nine goons and used the presence of one of these players in the box score for the team or the opponent as a pair of indicator variables.\(^{10}\)

The final control variables account for the recent tendency for a team to be called for penalties and the recent tendency of the opponent to draw penalties. One variable is the 15-game moving average for the number of times each game the team has been shorthanded. The other variable is the 15-game moving average for the number of times the opposing team has garnered a power play opportunity.

To isolate the deterrent effect, we need to find instruments that are correlated with the number of referees assigned but are not correlated with the error term, \( \epsilon \). We take advantage of the fact that all teams were to play the same number of games with the extra referee over the course of the season, but the distribution of the two referee games was randomly determined (until March 22, 2000, which ends our sample period) prior to the start of the season. This suggests that past history of referee assignment should now be a reasonable predictor for the current game. The more games each team has already played with only a single referee, the more likely they are to be assigned a second referee for the current game. As discussed, the 1998–99 experiment should have allowed teams to become familiar with the two-referee system, so the number of penalties called in the current game should not be correlated with the number of previous single referee games. We therefore construct two variables to serve as instruments: the percentage of the home team’s previous games that had only one referee, and the percentage of the opponent’s games that had only one referee.

With any IV regression, there is the risk that the instruments will be weak, and therefore a true underlying effect will be masked. The evidence suggests that our instruments do not suffer from this problem. In particular, they pass the \( TXR^2 > 2 \) and \( p^2 > 1/T \) tests suggested by Nelson and Startz (1990) for determining instrument relevance.\(^{11}\) Our computed statistics are 36.04 and and 22.60, respectively. In addition, as suggested by Bound et al. (1995), we note the \( R^2 \) from regressing the number of referees on all the exogenous variables climbs from 0.0083 to 0.056 (a 577% increase) when the two instruments are added, with \( t \)-ratios of 5.11 for the home team instrument and 3.09 for the opponent team instrument.

IV regressions are presented in Table 3. The number of referees is not statistically significant in any of the regressions, suggesting that players do not commit fewer infractions in response to the increased number of referees. This also suggests the simple difference in means reported in Table 2 represents a strict monitoring effect, not dampened to any significant degree by deterrence. Thus, we conclude that an additional referee catches infractions that otherwise might be missed, but the players themselves do not take this into consideration.

Of the benefit and cost variables, the team save percentage is statistically significant for all regressions except for majors. The opponent’s shooting percentage is statistically significant only for majors. Both of these results have the expected sign. Coefficients for the recent tendency variables have the predicted sign and are generally significant. As expected, teams typically having more power play opportunities are better at drawing (minor) infractions from the other team. The goon variables are generally significant. We also find that older players tend to commit more infractions, in particular more major infractions.\(^{12}\) To a lesser degree of confidence, we also conclude that height and weight differences impact the number of minor infractions committed, and a team’s power play defense may contribute to its likelihood of committing a major infraction.

VI. CONCLUSION

In our study of the NHL’s experiment, we find a statistically significant increase in the number of penalties assessed in games with

\(^{10}\) The goons were the 10 players who accumulated the most penalty minutes during the year, minus 1 player who had fewer minutes than another teammate.

\(^{11}\) \( p^2 \) is defined as \( \frac{[(T - 1)R^2 - 2]}{(T - 2)} \).

\(^{12}\) An alternative interpretation is that other players on teams that have older players commit more majors, perhaps to protect the older players. Using team averages, we cannot specify which particular players on the team commit the majors. Ecological fallacy arguments may apply here.
an extra referee, suggesting a strong monitoring effect took place. We utilize an IV routine to estimate the impact of an extra referee on the number of infractions. We do not find a significant deterrent effect, which implies players do not consider the number of referees in the game an important determinant for committing infractions. Because many sports infractions take place during the heat of competition and may be accidental or retaliatory in nature rather than planned in advance, the act of committing a sports infraction may be more analogous to a crime of passion rather than a calculated benefit-cost analysis performed by a rational criminal.

REFERENCES


Cairns, J., N. Jennett, and P. J. Sloane. ”The Economics of Professional Team Sports: A Survey of Theory and...


