



Do salaries improve worker performance?

Alex Bryson^{a,*}, Babatunde Buraimo^b, Rob Simmons^c

^a National Institute of Economic and Social Research and Centre for Economic Performance, LSE, United Kingdom

^b University of Central Lancashire, United Kingdom

^c Lancaster University Management School, United Kingdom

ARTICLE INFO

Article history:

Received 14 April 2010

Received in revised form 9 December 2010

Accepted 13 December 2010

Available online 21 December 2010

Keywords:

Incentives

Salaries

Productivity

Sports

ABSTRACT

We establish the effects of salaries on worker performance by exploiting a natural experiment in which some workers in a particular occupation (football referees) switch from short-term contracts to salaried contracts. Worker performance improves among those who move onto salaried contracts relative to those who do not. The finding is robust to the introduction of worker fixed effects indicating that it is not driven by better workers being awarded salary contracts. Nor is it sensitive to workers sorting into or out of the profession. Improved performance could arise from the additional effort workers exert due to career concerns, the higher income associated with career contracts (an efficiency wage effect) or improvements in worker quality arising from off-the-job training which accompanies the salaried contracts.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Although psychologists point to the inherent value of work to human beings and identify the beneficial impact it has on their wellbeing, the rational economic behavioural model underpinning labour economics equates effort with disutility. As such, economists assume employees are motivated by monetary rewards. Furthermore, they can be induced to expend greater effort in a task if those efforts are rewarded directly through performance-related pay. Although performance-related pay contracts are increasingly common, most workers receive an annual salary. If some of their income is tied to performance, it usually constitutes only a relatively small amount of total remuneration. One reason for the preponderance of salaried contracts is that workers are risk-averse and may be disinclined to share the risk of poor performance with their employer. Salaried employment minimises fluctuations in pay, thus giving workers some degree of certainty about their income into the future, provided they reach the performance threshold necessary to retain their job and provided the employer remains in business.

Where workers are heterogeneous they may choose between contract types according to their risk preferences and their ability. Employers wishing to share risk with workers may be willing to pay a wage premium for performance-related contracts. More able workers may seek performance-related contracts in the expectation that this will provide them with a higher income than a standard contract

(Lazear, 1986; Prendergast, 1999: 14–15). However, workers in a particular profession or occupation are not normally free to choose between contract types, either because a profession is characterised by a particular type of contract or because, if there are various types of contract on offer, the employer may be the arbiter of who manages to enter the jobs offering the “better” contracts. In this case employers can be expected to offer those job slots to the more talented workers.

The empirical literature has made only limited progress in establishing the incentive effects of various contract types. This is due primarily to data limitations. The bulk of the literature presents quasi-experimental estimates of switches in payment regimes at the level of the firm, often based on personnel data from a single case study firm, or a set of firms in a particular industry. We contribute to the literature by exploiting panel data on individual workers in an occupation where some but not all of the contracts on offer switched from very short-term contracts to salaried contracts providing us with the basis for a difference-in-difference estimation strategy. We observe workers over time pre- and post- the switch in payment regime which means we can account for fixed unobservable worker attributes – such as ability – which may affect workers' probability of being offered a salaried contract. Because we observe workers over time leaving and entering the profession we are also able to examine the extent to which any performance effects of a switch to salaried contracts is associated with worker selection as opposed to direct incentive effects.

Our paper explores the effects of a switch away from match fees towards salaried contracts among football referees officiating at English Premier League games. We find that cards per game – our measure of referee performance – fell for salaried referees after 2001,

* Corresponding author. Tel.: +44 7969179755.

E-mail address: a.bryson@niesr.ac.uk (A. Bryson).

after controlling for confounding influences such as attendance, team standings, match score and season. The paper considers some possible reasons for this finding. Section 2 reviews the evidence relating to the performance effects of individual performance-related pay and salaries. Section 3 describes our institutional setting. Section 4 presents a model of referee behaviour linking compensation schemes to referee performance. Section 5 describes the data. Section 6 presents our estimation strategy. Section 7 presents results while Section 8 concludes.

2. Previous literature

Performance-related-pay can improve worker performance through direct incentive effects as workers expend more effort creating the outputs that are rewarded through pay, and through worker sorting since more able workers have more to gain from a pay system which rewards them according to their performance (Lazear, 1986). Firm case studies show worker productivity rises with switches to piece rate payments with worker sorting effects often accounting for a sizeable part of the total impact. The productivity of window-screen fitters rose by 44% when the company Safelite switched to piece rate payments, with worker sorting accounting for half the increase (Lazear, 2000). Shearer (2004) obtains similar results using an experimental design whereby tree planters were randomly allocated to a piece rate and fixed rate. Bandiera et al. (2005) show productivity of fruit pickers rises dramatically with a shift from individual incentive pay based on relative performance (fruit pickers' mean pay set ex ante for the field but divided up according to relative performance) to piece rate (pay per unit of output fixed ex ante that does not vary with co-worker performance). In a companion paper Bandiera et al. (2007) emphasise the role managerial performance bonuses can play in improving worker productivity, an effect that comes via managers focusing their efforts on the most able workers and in recruiting and retaining better workers. Two papers estimate the effects of a move from piece rate to salaries. Fernie and Metcalf (1999) found that horse jockeys who switched to a salary-like retainer performed more poorly than jockeys who continued to be paid a piece rate, although their comparison is based on a small sample size. Similarly, Freeman and Kleiner (2005) show labour productivity in a shoe manufacturer fell following a shift from piece rate to time rates.

Despite these positive productivity findings, piece rate pay is in decline whereas performance-related pay more generally is on the rise (Lemieux et al., 2009). This is for a number of reasons. First, there is a growing realisation that workers paid by the piece will focus exclusively on those activities which trigger payments, often to the detriment of other activities which an employer might wish to encourage. This has led to the increased usage of subjective as opposed to objective assessments of worker performance by supervisors and management which allow the employer to make a more rounded appraisal of the worker's achievements. Second, workers can "game" piece rate systems. For instance, they may collude in reducing effort during a period in which the piece rate is set, thus lowering the output thresholds needed to trigger additional payments. Third, since piece rates require the worker to share the risk of underperformance with the employer, the employer needs to pay a premium to encourage risk-averse workers to accept piece rate payments. Indeed, it is often unclear in empirical work how much of the performance enhancement attributed to the introduction of piece rate working is actually an income effect akin to an efficiency wage, rather than the linking of pay to performance per se. Fourth, the growth in team-based production has increased the popularity of group-based performance-related pay ("gain-sharing") and financial participation based on profit-sharing or share ownership (Pendleton et al., 2009). Fifth, piece rate pay can entail increased costs such as additional worker compensation and monitoring which can more than offset

productivity gains, resulting in lower profitability. This is precisely what Freeman and Kleiner (2005) found in their study.

Most workers are on salaried contracts. They have their own incentive effects. As Prendergast notes (1999: 10), where a worker is paid a fixed salary in a given period "despite the fact that there is no immediate relation between pay and performance, he is likely to have incentives to exert effort because good performance will improve future contracts. Such reputational concerns imply that effort exertion can occur without explicit pay-for-performance contracts." In long-term salary contracts, the benefits of good performance may come in the form of deferred payments or benefits, such as pensions. In shorter fixed-term salary contracts, the rewards to good performance may come in the form of contract renewal. Salaries also smooth income fluctuations for workers, offering income security which may be absent among those on shorter-term performance-oriented contracts. These incentives may be particularly powerful if salaried job slots are rationed, creating a tournament among competing workers for those salaried contracts. In such cases, one may anticipate more able workers competing for the salaried contracts.

3. Institutional setting

Prior to the 2001/02 season all football referees in England were paid a match fee as compensation for officiating (Rickman and Witt, 2008). The fee was a flat rate but was higher in the top league – the Premier League, which we will call "tier 1" – than the second league – known as the Championship, which we label "tier 2" games.¹ Since the 2001/02 season referees officiating at tier 1 games were paid a salary in a contract which was renewable at the end of two seasons (more recently, the end of each season) and which also included the match fee for officiating at a game.

Rickman and Witt say the motivation for salaried contracts paid to referees was the need for 'a clearer regulatory regime, consistent training, development and monitoring regimes' (2008: 299).² The Premier League had already established a monitoring system by which referees' performances was assessed by peer review. But the League experienced large growth in revenues, fuelled by lucrative sales of broadcast rights, and greater media exposure, part of which focussed on key decisions made by referees in high-profile games. This encouraged the Premier League to offer salaried contracts to a Select List of referees. A company was set up (the Professional Game Match Officials Board, PGMOB) to provide match officials for all professional games played in the English Leagues (Premier League and Football League). The PGMOB is jointly funded by the English Football Association (the governing body of football in England), the Premier League and the Football League.

Table 1, which uses the data described in Section 4, shows the numbers of referees officiating in the top two divisions in English League football since 1997/98. The total number of referees is roughly constant over the period at around 70. Of these, roughly 50 officiated in any given season in tier 2 and around 20 officiated in tier 1. Since the 1999/00 season those officiating in tier 1 also tended to officiate in tier 2. Prior to this those officiating tier 1 matches tended not to officiate at tier 2 matches. The final column shows the number on salaried contracts after 2001. Initially 24 contracts were awarded but in most seasons since then the number awarded has been closer to 20.

Table 2 shows the mean numbers of games officiated by referees, by tier and after 2001 by contract type. Match fee referees average 8–9

¹ Tier 3 games attracted a lower rate still and the system of rising match fees as League status increases remains in force today.

² A further argument in favour of salaried contracts for referees has surfaced in recent years with the growth of betting activity, especially via online sources. By raising the salaries of referees at the top level, any propensity they may have to become involved in match-fixing will be reduced as the expected gain from corrupt activity will be less and the incidence of match-fixing will be reduced (Forrest et al., 2008).

Table 1
Number of referees who officiated in both pre- and contract periods.

Season	Tier 1 only	Tier 2 only	Tiers 1 and 2	All referees	Referees on salaried contracts
1997–98	19	49	0	68	0
1998–99	17	53	1	71	0
1999–00	0	50	21	71	0
2000–01	3	52	23	78	0
2001–02	1	51	24	76	24
2002–03	1	51	22	74	25
2003–04	1	55	21	77	21
2004–05	0	50	23	73	19
2005–06	0	40	22	62	16
2006–07	0	47	20	67	17
2007–08	0	49	18	67	18
2008–09	1	51	18	70	19

Table 2
Mean number of games per season by tier and type of contract.

Season	Tier 1		Tier 2	
	Contract	Non-contract	Contract	Non-contract
1997–98		20.00		10.98
1998–99		21.11		10.22
1999–00		18.10		7.76
2000–01		14.62		7.36
2001–02	15.79	1.00	6.58	7.73
2002–03	16.52	0.00	5.88	8.39
2003–04	18.05	1.00	5.52	7.93
2004–05	19.42	2.75	5.00	8.46
2005–06	22.94	2.17	6.50	9.74
2006–07	22.06	1.67	7.12	8.62
2007–08	21.11	0.00	7.33	8.57
2008–09	20.00	0.00	6.17	8.65

games per season. In the first years of salaried contracts those on salaries tended to officiate at 16–18 tier 1 matches and 5–6 tier 2 matches. But from around 2005/06 they were officiating at 20–22 tier 1 matches and 6–7 tier 2 matches. Thus the total number of games that salaried referees officiated per season rose over the period.

One institutional feature of top-level refereeing that persisted initially in spite of the introduction of salaried contracts was a compulsory retirement age of 48, with a small number of exceptions. A concern of the football authorities and other stakeholders in the game is that referees who are ‘too old’ may lack the physical fitness needed to keep with the fast flow of play in top-level professional football.³ This age restriction has recently been abolished, in line with European Union law, although Select List referees must still pass rigorous fitness tests.

As part of the professionalization of football refereeing the salaried contracts stipulated that the referee had to undertake off-the-job training on a regular basis to improve his or her skills in officiating at a football game. These fortnightly training sessions combine rigorous and specialist fitness training with reviews of previous performance using video playback. The training sessions are held at Premier League

³ Older age might be associated with improved experience, better ‘reading’ of the game and with referees taking up better positions on the pitch so as to make good decisions. Weston et al. (2010) analyzed a panel data set of Prozone statistics detailing aspects of physical performance by referees in Premier League games. They find that older referees (aged 43 to 48) had lower total distance covered, less average running distance at high intensity speed and fewer sprints compared to younger referees (aged 31 to 36). Although physical performance appeared to decline with age, the authors find that this did not impact upon the ability of older referees to keep up with play since average distance from location of fouls was not significantly different between the two age groups. This suggests that learning and experience are important features of top-level refereeing.

Table 3
Salaries and tier 1 match fees for referees after 2001.

Season	Salary	Match fee
2001/02	33,000	600
2005/06	50,000	230
2007/08	53,000	270
2008/09	57,000	350

Sources: Sawdon-Smith (2001) for 2001/02, www.independent.co.uk/news/uk/this-britain/what-britain-earns-509669.html for 2005/06, www.timesonline.co.uk/tol/sport/football/premier_league/article2296121.ece for 2007/08, www.guardian.co.uk/football/2008/aug/14/premierleague1 for 2008/09.

clubs so that, to some extent, referees actually train with players. This was intended to give referees greater insight into player (mis)behaviour and to facilitate better communication between referees and players. On the other hand, there is also scope for regulatory capture.

The extra training sessions meant that salaried referees had longer hours than match fee referees who only needed to officiate on match day. Rickman and Witt (2008) report that in 2001/02, when salaried contracts were first introduced, salaried referees received an annual retainer fee of £33,000 per annum.⁴ From 2001/02 they also received a match fee of £600 per game (Sawdon-Smith, 2001). The figure of £600 was actually a reduction from £900 in the previous season.⁵ By 2008/09, the basic annual salary figure of £57,000 (James, 2008). In most cases this meant that salaried referees could focus on refereeing as their primary occupation without having to rely on second jobs to supplement their income. Indeed, in 2001/02, there was only one referee who rejected the offer of a salaried contract in order to retain his regular job.⁶

Table 3 reports estimates of contracted referee salaries and tier 1 match fees, obtained from newspaper sources. It is clear that, in both nominal and real terms, salaries for referees appointed to the Select List have grown considerably since 2001/02.⁷ For 2005/06, salaries were £50,000 and the tier 1 match fee was reported to be £230. Using the mean games from Table 2, this implies an annual gross income of £55,276. This compares with the non-contract average of £2,739. For salaried referees in 2008, an annual retainer of £57,000 plus match fee of £375 (multiplied by mean number of tier 1 games of 20) yielded £64,500. These referees also officiated an average of 6.17 games in tier 2 (at £145) giving an additional £895. This suggests an estimated annual gross income of £65,395 on average.⁸ Non-contract referees in the same season would have received $8.65 \times £145 = £1,254$ on average, a gap of £64,141. Note that the match fee for non-contract referees has gone down between these years. The earnings gap between contracted and non-contract referees is widening over time if the sources of revenue are limited to tiers 1 and 2. Non-contract referees will also earn money from officiating in tiers 3 and 4 but the

⁴ The football authorities were able to afford this because the contracts were introduced at a time when more money was flowing into the game from lucrative television rights (Dobson and Goddard, 2001).

⁵ The figures of £600 and £900 included expenses for travel, accommodation and subsistence. After 2001, these expenses were awarded separately on top of a lower match fee. This accounts for the lower match fees reported in Table 3 after 2001.

⁶ This was David Elleray who retained his position as house master at Harrow School but continued to referee Premier League games, paid by match fee alone. To be able to do this he had to gain special dispensation from the PGMOB to be excused from attending prescribed training and development sessions. Other referees took up jobs, such as taxi driving, with flexible hours that could be accommodated around a busy schedule of games during the football season (Sawdon-Smith, 2001). Self-employment remains an option for salaried referees.

⁷ This has led to complaints by referees’ assistants (formerly linesmen) and some Football League referees not on the select list that their match fees have not risen in line with contracted referee salaries. These complaints have not been upheld by the PGMOB. See www.guardian.co.uk/football/2008/aug/14/premierleague1.

⁸ This does not include income from Cup games, many of which are officiated by salaried referees in the final stages.

earnings gap with ‘professional’ referees remains substantial and increasing over time.

In England, poorly performing referees on retainer can be, and sometimes are, demoted to lower status Football League games. When select contracts were first introduced in England in 2001/02, the duration of contract was fixed at two years but from 2003/04 select referees were awarded one year rolling contracts, making it easier for the PGMOB to fire poorly performing referees and promote promising referees from the National List.

This discussion suggests there are at least four mechanisms by which salaried contracts might have elicited improved referees’ performance relative to the match fee system: the incentive effects arising from career concerns; the efficiency wage effect associated with higher wages; the ‘quality’ effect associated with off-the-job referee training; and worker sorting leading to better workers obtaining salaried contracts. If salaried contracts were superior to match fee contracts, one might expect more able referees to queue for the relatively small number of core contracts offered by the football authorities. Since there are normally ten tier 1 games to officiate each weekend there appear to be sufficient salaried referees to undertake the task throughout the period. However, the number of salaried contracts gradually fell throughout the period from 24 in 2001/02 to 19 in 2008/09 (Table 1), implying increasing competition for these slots.

All tier 1 games were officiated by referees on salaried contracts after 2001/02 with a few exceptions which are discussed below. Referees not on a salaried contract continued to be paid on a per match basis after 2001. However, all of them were eligible in principle for salaried contracts, if they were promoted to tier 1 games. Thus the competition for salaried posts is akin to a tournament and may have generated improved performance among both salaried and non-salaried workers keen to obtain a salaried position in the following season. In Section 6 below we explain how we exploit this natural experiment to investigate whether salaried referees performed better than those on match fees.

As discussed in Section 4, refereeing is a complex activity requiring the worker to perform multiple tasks which is why referees are assessed by the football authorities using subjective assessments by expert observers. A good referee is one that keeps the flow of the game going and ensures that fair play is enforced. Good refereeing relies on having experience, the right temperament, fitness and the ability to work in a team with the other officials (two assistant referees, one on each touchline, and the fourth official who assists with matters such as ensuring managers stay within their designated areas). Referees’ performances are carefully scrutinised by paying fans and by the media who, with the benefit of instant television replays relayed around the ground, can assess whether key decisions were correct or not. A considerable body of literature finds evidence that referees’ decisions are influenced by the social pressures fans exert.⁹ We are able to account for this with controls for game characteristics including matches between local rivals and crowd size. The fact that referees’ performances are readily evaluated by fans, the media and the football authorities means that the cost of a poor performance to a referee is likely to be substantial and immediate. Those on a match fee may find they are not selected for next week’s match while those on a

salary will suffer a blow to their reputation and may find it harder to have their salary contract renewed in the next season. A further implication is that there is limited room for referees to “game” the system of incentives they are working under: the efforts (whether they keep up with the play, make the right decisions, keep the game flowing and keep the players under control) are observable to some extent. Nevertheless, there is enough discretion available to referees in their interpretations of the Laws of the Game to mean that principal-agent problems cannot be ruled out (Rickman and Witt, 2008).

4. A model of referee behaviour

Let the production of value from refereeing be $V = f(C, P)$. C is effort devoted to card giving which we shall term ‘deterministic’ effort while P is preventive effort. We assume that $f_C > 0, f_P > 0, f_{CC} < 0, f_{PP} < 0, f_{CP} > 0$ and $f_{CC}f_{PP} > f_{CP}^2$ for concavity of the value function. The cost of effort to the referee is $X = g(C, P)$ where $g_C > 0, g_P > 0, g_{CC} > 0, g_{PP} > 0, g_{CP} > 0$ and $g_{CC}g_{PP} > g_{CP}^2$ for convexity. P and C are substitutes in the production of value and in the referee’s cost function. Suppose the referee receives a fixed salary that is not conditional on effort, that C is observable and verifiable and that P is exogenous then a single League’s surplus is:

$$V - W - X \text{ where } W \text{ is a fixed fee paid to referees.} \tag{1}$$

Assume that the League can determine C while P is exogenous. C is given by the number of cards awarded by referees. The League planner maximises its surplus by varying C .

$$f_C - g_C = 0 \tag{2}$$

and the second order condition for a maximum surplus is $f_{CC} - g_{CC} < 0$ which follows directly from the concavity and convexity conditions stated above. If P varies exogenously, what is the first-best optimal response of C for the League planner? Implicit differentiation of Eq. (1) and using the 1st order condition (2) gives:

$$dC / dP = (f_{CP} - g_{CP}) / (g_{CC} - f_{CC}) \tag{3}$$

$dC/dP < 0$ follows if $g_{CP} > f_{CP}$ so the impact of an extra unit of P on the marginal value product of sanctions is less than the corresponding impact on the marginal cost of effort with respect to card giving. Hence, under these assumptions, there is an inverse relationship between P and C at the optimum. If P rises then C must fall to preserve an optimum level of surplus for the League planner.

Now suppose both C and P are variable. Let φ be a weight on P in the value function so this becomes $f(C, \varphi P)$. Then in the first-best solution:

$$\begin{aligned} f_C - g_C &= 0 \\ \varphi f_P - g_P &= 0 \end{aligned} \tag{4}$$

Suppose the weight on P in the value function increases for some reason. Then comparative static analysis shows that $dC/d\varphi < 0$ and $dP/d\varphi > 0$. C declines and P goes up.

A greater weight on P in the value function is precisely what has occurred in the English Premier League (tier 1) since its inception in 1992. The increased rewards at the top level of English football, stimulated by enhanced revenues from broadcast rights, have encouraged greater athleticism and faster speed amongst the players appearing at this level. As the game has become faster, the marginal product of P in the production function has increased. To see this, consider a CES production function;

$$L = [(A_P P)^\sigma + (A_C C)^\sigma]^{1/\sigma} \text{ where } 0 < \sigma \leq 1. \tag{5}$$

⁹ See Garicano et al. (2005), Rickman and Witt (2008) and Dohmen (2008) on bias towards home teams in the form of time added on at the end of games at the referee’s discretion. Dawson et al. (2007) and Buraimo et al. (2010) present evidence of bias towards home teams and against away teams in the award of sanctions (yellow and red cards). Dohmen (2008) suggests that referee decisions such as penalty awards may also contain non-random mistakes. This evidence covers a number of dimensions of referee decision-making across several professional football Leagues, principally England, Germany and Spain, and points towards systematic biases in referee behaviour. Rickman and Witt (2008) present evidence to show that biases associated with discretionary time added on by referees were eliminated after salaried contracts were introduced to the English Premier League.

The log of the marginal rate of transformation between P and C is given by:

$$\ln MRT_{C,P} = \sigma \ln(A_P/A_C) + (1-\sigma) \ln(C/P). \quad (6)$$

The increased weight on P in the production of value leads to an increase in the preventative effort component of referee activity in a way that is analogous to the increased demand for skilled labour due to skill-biased technological change.

Referees perform two ‘tasks’ (card giving and preventive effort) but there is just one source of reward that is contingent on the two efforts associated with these tasks. It is not possible for the principal to offer separate rewards for distinct tasks and, even if it were possible, P is difficult for the principal to observe while C , although easier to observe, is a noisy measure of sanctioning effort since the card giving process is partly driven by a stochastic process which depends on factors such as the good behaviour of players. If this was not the case, the employer could simply make pay contingent on observed effort. Even if cards were a suitable measure of effort devoted to sanctions, a contract based purely on card giving would fail to consider the impact of discretionary effort on value. Thus the League needs to avoid a corner solution where referees only devote effort to cards and perform zero preventive effort. A corner solution is ruled out by the assumption of different productivities of the two types of effort. A corner solution is optimal if card giving is less costly to the referee at the margin than preventive effort but the marginal products of each type of effort are equal or the marginal product of card giving is greater than the marginal product of discretionary effort. From Eq. (4) if $g_P > g_C$ then at the first-best optimum it must follow that the marginal value product of discretionary effort φ_P is greater than f_C for the optimum to be at an interior solution.

Efforts are not easily observable but assume combined effort has a performance metric attached to it, which we denote by x . This performance metric is observed by the governing body but is not verifiable. In the empirical analysis below, we proxy the non-revealed (to the econometrician) values of x by performance ratings taken from newspaper match reports. The efforts in each of the two tasks determine performance as follows:

$$\Pr(x = 1 | e) = \eta C + P + \varepsilon \quad (7)$$

where $\eta > 1$ and ε is an i.i.d error term which partly reflects the stochastic elements in card giving.¹⁰ Our framework is one where refereeing ‘output’ (performance) is observable but stochastic. The principal sees if a referee’s performance is good but cannot tell whether this is due to spontaneous good behaviour of the players or to good refereeing. Similarly, card giving is a noisy measure of effort, C , as sanctions will result from both the spontaneous misbehaviour of players and poor control of the game by a referee. For this reason, we do not observe card-contingent contracts.

The parameter η in Eq. (7) illustrates the multi-tasking problem in our model. If $\eta > 1$, a referee who is rewarded on the basis of x has incentives to substitute away from P and devote more effort to C . Note that this substitution would occur even if each type of effort had the same marginal effect on the referee’s cost of effort and the same marginal effect on the League’s value. The substitutability of efforts in the referee’s cost function merely reinforces the result that holds with an additively separable effort function. Moreover, the multi-tasking problem is exacerbated if the two types of effort are substitutes in the production function with differences in their marginal products.¹¹

¹⁰ Our analysis of optimal referee contracts follows Corts (2007) and Mukherjee and Vasconcelos (2010), who develop models of optimal job design with implicit incentive contracts.

¹¹ This problem, sometimes referred to as “teaching to the test”, is discussed by Holmstrom and Milgrom (1991: 32–33) and Lazear (2006).

Since performance in each job (League) is non-verifiable, the League can only offer an *implicit contract* promising a *bonus* payment if referee performance is found to be good. The contract offered to referee A is a pair (W, B) where W is a lump sum wage and B is an implicitly contracted bonus payment offered only if performance is good, that is, $x = 1$. We can relate these payments to what we actually observe in the world of football refereeing as follows. In a given season there are n games to be refereed in the League. There are fewer qualified referees available, N , than total games to be refereed so some rationing must occur. If we assume identical referees, then each referee will officiate n/N games. Suppose the League governing body offers a per game fee of F for League games then the fixed payment for a typical referee in each League will be Fn/N . The bonus B can take two forms. First, it can entail a redistribution of games such that a referee posting good performance is assigned to more games than one who delivers poor performance so referees do not officiate equal numbers of games. Second, the principal can reward good performance by an increment to the expected payment from a given number of games. Both of these features are observed in practice. The crucial points about this bonus are first, that it is promised, and second, that the promise is contingent on satisfactory effort.

The annual salary, as opposed to a per match fee, creates career concerns for selected referees but the credibility condition requires that the total remuneration to these referees should increase.

The League also instituted a two-tier structure for referee remuneration. In the top tier, the Premier League, there was a salary per annum. In the second tier, the Football League, referees were compensated by a basic fee per game. The expected annual remuneration for refereeing in the Premier League was greater than for the Football League. For Premier League referees this means that the outside option increased to being the expected earnings from an alternative occupation plus earnings from refereeing in the Football League. This in turn raised the opportunity cost of a bad performance for Premier League referees.

Although credibility may be satisfied (in the sense that the League does not renege on its promised bonus), the incentive compatibility condition – for optimal referees’ effort choices to be interior solutions – may still not be satisfied. Returning to Eq. (7), the League may assess referee performance with a new weight of $\eta < 1$ so as to reflect the increased weight of preventive effort in the value function. Referees then have an incentive to substitute preventive effort for card giving effort. With $\eta < 1$, however, incentive compatibility is violated by $X_P > X_C$ in the cost of effort function, which is the assumption we made at the beginning of this section. To see this point, consider the payoffs of a representative referee defined as:

$$U = W + B \Pr(x = 1 | e) - X(C, P) \quad (8)$$

The outside option of the referee is assumed to be zero. The incentive compatibility constraint requires maximisation of the referee’s net payoff to yield the first-order conditions:

$$\begin{aligned} B\eta - X_P &= 0 \text{ and} \\ B - X_C &= 0. \end{aligned} \quad (9)$$

Hence, if $\eta > 1$ then X_P must exceed X_C for the incentive compatibility conditions to hold. Conversely, if X_P does exceed X_C , as we propose was actually the case, then $\eta < 1$ will violate the incentive compatibility conditions. The League offered a rational response to this incentive compatibility problem: by offering, indeed making compulsory, appropriate training and educational support to referees involving more extensive performance reviews and better communication with teams and players prior to match day, the League was able to help referees reduce the marginal cost of preventive effort to the referee relative to the less variable marginal cost of card

giving.¹² However, this off-the-job training and education process requires additional effort on the part of a referee and the salary payment, made conditional on undertaking training and educational development, is needed to sustain incentives to participate in this function.

5. Data

There were 11,184 tier 1 and tier 2 league games played in England in twelve seasons between 1997/98 and 2008/09. We have referee information on all but 15 of these games. Forty-one percent of the games in our data are tier 1 games. Salaried contracts were introduced in the 2001/02 season for referees officiating in tier 1 games. Across the whole period, two-thirds of tier 1 games are officiated by salaried referees.

All tier 1 games were officiated by referees on salaried contracts after 2001/02 with a few exceptions which are discussed below. However, some of those with salaried contracts officiated in tier 2 games after 2001/02: these matches accounted for 15% of all tier 2 games.

We know the identity of all 168 referees who officiated in these games and this represents the whole population of referees. Of these referees, nine only ever officiated in tier 1 games, 114 only ever officiated in tier 2 games and the remaining 45 officiated in both tier 1 and tier 2 games. The fact that we find a sizeable percentage of referees who officiate at both levels indicates that the referees' labour market is not segmented by tier, thus lending credence to our suggestion that all referees employed at the time salaried contracts were introduced could, in principle, be eligible for such a contract.

At the referee level we have an unbalanced panel. They officiate at between one and 312 games, the mean being 66 games. Sixty-nine referees are observed in the period before and after the introduction of salaried contracts, including 37 referees who had a salaried contract at some point after 2001/02.

We measure performance with the number of yellow and red cards referees issue in a game. Yellow cards are issued as a warning to a player when he has broken the rules, either by fouling an opponent, handling the ball or for showing dissent. A red card is shown if the player commits a second offence worthy of a yellow card. It can also be shown for a first offence if the offence merits it, such as violent conduct or in the case of a foul which directly prevents a goal scoring opportunity. Red cards lead to the player being sent off and suspended from subsequent games. Issuing many cards is often a sign that a referee has lost control of a game. Good referees are able to deal with most incidents without brandishing cards by communicating firmly with players from the outset. Indeed, it appears that the Premier League saw improved communication between referees and players as an important outcome of the conditions of salaried contracts.

Ideally, we would observe overall referee effort. We would like to use peer reviews of referees undertaken by the referees' employer but these are proprietary. The closest indicator one can obtain is subjective assessments of referee performance by one newspaper, The Times.¹³ Another study from Germany has shown a high negative correlation between number of cards awarded and magazine assessments (Frick et al., 2008). We have coded referee ratings on

¹² Training and development programmes can lower marginal costs of each type of effort, but we contend that the marginal cost of preventive effort fell by more than the marginal cost of effort devoted to sanctions when these programmes were introduced by the referees' employer.

¹³ Inspection of the content of match reports associated with the publication of referee ratings and conversations with referee peer assessors suggest to us that the criteria for good and bad performances of referees are consistent between journalists and assessors. Features such as allowing the game to flow, getting key decisions right and generally showing good control without being overly intrusive are indicative of good performances as judged by both assessors and journalists. See Appendix A for further details on journalists' assessments of referees.

Table 4
Summary statistics.

	N	Mean	S.D.	Min	Max
<i>Panel A: All games</i>					
Yellow cards per game	10,904	2.977	1.915	0	12
Second yellow and red cards per game	10,904	0.174	0.441	0	4
TIER 1	10,904	0.407	0.491	0	1
POST00/01	10,904	0.668	0.471	0	1
TIER 1 × POST00/01	10,904	0.271	0.445	0	1
Attendance	10,904	23,106.120	12,808.830	849	76,098
SCORE	10,904	0.389	1.650	-7	8
SCORESQ	10,904	2.874	4.765	0	64
DERBY	10,904	0.045	0.207	0	1
MONTHSTOGO	10,904	4.609	2.669	0	9
HOMEPOINTS	10,904	1.348	0.493	0	3
AWAYPOINTS	10,904	1.380	0.497	0	3
<i>Panel B: Games up to and including 2000/01 season</i>					
Yellow cards per game	3625	2.963	1.985	0	12
Second yellow and red cards per game	3625	0.164	0.444	0	3
TIER 1	3625	0.408	0.492	0	1
POST00/01	3625	0.000	0.000	0	0
TIER 1 × POST00/01	3625	0.000	0.000	0	0
Attendance	3625	21,058.960	11,974.820	3436	67637
SCORE	3625	0.439	1.691	-7	8
SCORESQ	3625	3.052	5.219	0	64
DERBY	3625	0.047	0.212	0	1
MONTHSTOGO	3625	4.568	2.712	0	9
HOMEPOINTS	3625	1.350	0.488	0	3
AWAYPOINTS	3625	1.378	0.493	0	3
<i>Panel C: Games after 2000/01 season</i>					
Yellow cards per game	7279	2.984	1.879	0	12
Second yellow and red cards per game	7279	0.180	0.440	0	4
TIER 1	7279	0.407	0.491	0	1
POST00/01	7279	1.000	0.000	1	1
TIER 1 × POST00/01	7279	0.407	0.491	0	1
Attendance	7279	24,125.620	13,086.410	849	76,098
SCORE	7279	0.364	1.629	-6	7
SCORESQ	7279	2.786	4.519	0	49
DERBY	7279	0.044	0.204	0	1
MONTHSTOGO	7279	4.629	2.648	0	9
HOMEPOINTS	7279	1.348	0.495	0	3
AWAYPOINTS	7279	1.381	0.499	0	3

1305 games in our data set for three seasons (2001/2, 2002/3 and 2003/4) for which ratings were published. Because we have no ratings for the period prior to the introduction of salaries we are unable to use the ratings to estimate the impact of salaries on referee performance. However, we are able to use the data to establish whether card giving is negatively correlated with referee ratings, as our theoretical model predicts. Multivariate analyses controlling for season dummies and referee fixed effects confirm that giving yellow and red cards is strongly negatively correlated with referee ratings.¹⁴

We will present several covariates as controls including a time trend, dummy variable for tier 1, match attendance, score difference between teams in a game, dummy variable for games involving local rivals, months left until the end of the season and the points the home and away teams had won up until the start of the game. The match-level means and standard deviations for all variables used in the analysis are presented in Table 4.

6. Estimation

The model presented in Section 4 showed that preventive effort is more productive for the employer than sanctioning effort (see Eq. (4)). We argued that the shift towards salaries, together with the

¹⁴ Full analyses are presented in the Appendix A.

attendant increased emphasis on training and development, would induce referees to focus more on preventive effort at the expense of sanctioning effort. Thus we anticipate card giving – the only measure of effort we have available for the pre- and post-periods – to fall with the switch to salaries. Our model of total yellow cards given in a particular match is:

$$\begin{aligned}
 Y = & \text{CONSTANT} + \beta_1 \text{TIME} + \beta_2 \text{TIER1} + \beta_3 (\text{TIER1} \times \text{POST01} / \text{02}) \\
 & + \beta_4 \text{LOGATTEND} + \beta_5 \text{SCORE} + \beta_6 \text{SCORESQ} + \beta_7 \text{DERBY} \\
 & + \beta_8 \text{MONTHSTOGO} + \beta_9 \text{HOMEPOINTS} + \beta_{10} \text{AWAYPOINTS} + \varepsilon.
 \end{aligned}
 \tag{10}$$

A similar model is estimated for total second yellow and straight red cards in a given match (R). We use a difference-in-difference estimator to identify whether the introduction of salaried contracts affected referee performance. Our unit of analysis is the football match and our dependent variable (our performance measure) is the number of cards the referee awards during the game. First we run an OLS model which includes a linear term capturing time (TIME), a dummy variable for a tier 1 match (TIER1) and an interaction between TIER1 and a dummy variable for the period from 2001/02 onwards which captures the introduction of salary contracts for referees officiating in tier 1 games ($\beta_3(\text{TIER1} \times \text{POST01}/\text{02})$). We run this model without controls and then test the sensitivity of the results to the inclusion of the controls described below.¹⁵ In an alternative specification we replace the linear time trend with year dummies to capture potential non-linear time effects on referees' performances. The year dummies are jointly significant and the model performs a little better than the one using a linear time trend so we report results with the year dummies.

The identifying assumption in these models is that the interaction between tier 1 games and the period after the introduction of salary contracts will capture the causal impact of salary contracts on referee performance. The assumption is that any significant interaction effect reflects the introduction of salary contracts, rather than other contemporaneous changes which might have differentially affected tier 1 referees relative to tier 2 referees. At this stage, we have 11,169 games for analysis.

The control variables help account for the fact that referee allocation to games may be non-random. It seems probable that those referees assessed as good by the football authorities are more likely to be appointed to officiate high-profile games than referees assessed as performing more poorly. Since high-profile games also have the potential to be more heated and thus more difficult to officiate, it is conceivable that some referees will be confronted with a more challenging set of match assignments than others, something which could bias the estimates of salary contracts on referee performance. We control for differences in the difficulty attached to particular match assignments with a range of match-level controls such as (log) attendance (LOGATTEND), a dummy variable for whether the game is a derby game between two nearby rivals (DERBY), and the league positions of the clubs playing the match, expressed as accumulated points per game at the time of the match, assessed separately for home and away teams (HOMEPOINTS , AWAYPOINTS). The number of cards awarded may depend on the closeness of game as captured by SCORE which is the goal difference between home and away teams in a given match and its square SCORESQ . This measure was used by Rickman and Witt (2008) in their study of the impact of

salaried contracts on discretionary time added on by referees. The need to allow sufficient games in the early part of a season to pass so as to accumulate sufficient information for HOMEPOINTS and AWAYPOINTS means that at this stage our sample size drops slightly, from 11,169 to 10,904.

The football authorities may be more constrained in being able to appoint a salaried referee to a tier 1 game in some years relative to others due to fluctuations in the number of salaried referee contracts in place. We therefore tested the sensitivity of our results to the introduction of this as a control variable. This made no difference to our key results so it is dropped from the models presented here.

One potential threat to this identification strategy is the likely non-random allocation of salary contracts to referees. It is probable that the football authorities offer salary contracts to the most talented referees who they choose from the queue of referees keen to make refereeing their sole or main occupation. If so, then being on a salaried contract may simply proxy referees' talent. Comparing the tier 1 match experience, as total games and average per season, for referees who made it to the Select List compared to those who were eligible but not appointed, we find that those referees who were appointed to the Select List had greater experience than those who were not appointed. This suggests that appointment to salaried contract status was primarily determined by human capital (experience) considerations.¹⁶ To tackle the selection issue in our match-level analyses we introduce referee fixed effects to our difference-in-difference estimator. These fixed effects control for fixed referee attributes such as talent. With their introduction the interaction effect captures the impact of the switch to salary contracts net of any pre-existing talent differential between salaried and non-salaried referees.

A further threat to our identification strategy is the unbalanced nature of our panel which reflects referee entry and exits. Referees' decisions as to whether to enter the profession or quit may, in part, be associated with the advent of salaried contracts. Indeed, as our earlier discussion indicated, employers often deploy what they hope might be better contracts in the hope that it induces worker sorting which is productivity-enhancing. We can establish how sensitive our results are to worker sorting by varying our estimation sample. Estimates run on games officiated by referees who contribute observations for the period prior to the introduction of salary contracts and the period after their introduction limits the impact referee sorting can have on the results. A comparison between these estimates and estimates using all observations in the sample will establish just how important worker sorting is in explaining any role played by salaried contracts. Models containing only those games officiated by referees contributing observations before and after the introduction of salaried contracts are likely to give us the cleanest estimate of the direct career incentive effects of salary contracts.

All the above analyses are conducted on different measures of the propensity to give yellow and red cards to see how robust the estimates are. We are still left with three possible direct incentive effects of salaried contracts: the income effect; the "pure" career concerns effect; and the off-the-job training effect. We explore the timing of effects in the empirical analysis to see if we can distinguish between these mechanisms.

¹⁵ It might be argued that number of cards is a count variable and so a Poisson regression estimator is appropriate. All models were estimated using Poisson regression and each model delivered similar qualitative results to OLS. We report OLS results for greater ease of interpretation.

¹⁶ A probit analysis of selection into salaried contract from the pool of eligible referees in 2001/02 supports this argument. The explanatory variables on the right-hand side of the equation which capture being a good referee – such as refereeing 'bigu games and the number of cards shown to players – are all lagged variables measured in the period prior to the introduction of salary contracts in tier 1. The coefficients on these variables are jointly insignificant, leaving just tier 1 match experience with the solitary significant coefficient. Of course, peer review of refereeing performances occurred before the advent of salaried contracts and we would expect the best referees to have the greatest top-level experience and then to go on to be appointed to the Select List.

Table 5
OLS and referee fixed effects models for N first yellow cards per game.

Variables	OLS	OLS	FE	FE
TIER 1	0.644*** (10.16)	0.460*** (6.25)	0.381*** (4.36)	0.233** (2.49)
POST00/01	0.305*** (3.24)	0.226** (2.39)	−0.022 (0.20)	−0.027 (0.27)
TIER 1 × POST00/01	−0.554*** (7.14)	−0.510*** (6.56)	−0.483*** (5.24)	−0.427*** (4.63)
LOGATTEND		0.220*** (4.82)		0.227*** (5.07)
SCORE		−0.024** (2.01)		−0.019* (1.68)
SCORESQ		−0.029*** (7.33)		−0.028*** (7.29)
DERBY		0.716*** (8.17)		0.724*** (8.44)
MONTHSTOGO		0.067*** (9.95)		0.067*** (10.21)
HOMEPOINTS		−0.094** (2.35)		−0.085** (2.17)
AWAYPOINTS		0.140*** (3.81)		0.157*** (4.36)
CONSTANT	2.620*** (38.55)	0.249 (0.59)	2.936*** (38.13)	0.453 (1.09)
Year Dummies	YES	YES	YES	YES
Observations	11,169	10,904	11,169	10,904
Adjusted R-squared	0.01	0.04	0.07	0.10

Absolute t statistics in parentheses.

- *** p<0.01.
- ** p<0.05.
- * p<0.1.

7. Results

Our results are match-level estimates of referee performance where the dependent variable is the number of cards given during the match. Table 1 shows only modest variation in the number of salaried contracts awarded in a season and not surprisingly the coefficient of *NCONTRACTS* was not significantly different from zero throughout. We therefore drop this variable from the reported estimates.

Table 5 presents estimates for the number of first yellow cards given per match for the whole period. We present models with year dummies: these are jointly statistically significant whereas the linear time trend does not have a significant coefficient. The interaction term for tier 1 matches and the period since the introduction of salaried contracts is negative and statistically significant in the OLS models in columns 1 and 2, indicating that the number of first yellow cards given fell by around a half a card per game with the shift to salaried contracts. The size of the effect only drops a little with the introduction of controls in model 2.

These control variables perform as anticipated. The number of cards shown is significantly higher in ‘big’ games, that is, those with high attendances and derby games between local rivals. However, if there is a large score discrepancy in the game – that is, the game proves not to be particularly competitive – this results in fewer cards being shown. This effect is non-linear, growing with larger score discrepancies. The number of cards is greater earlier on in the season, a finding that is consistent with players being more cautious towards the end of the season to ensure they are available for selection in important games. It is also consistent with the proposition that referees exert greater effort towards the end of the season, thus producing fewer cards, in the hope that this may increase their chances of having their salaried contract renewed or being offered one for the first time. Team performance prior to the game also proves to be a significant factor. Where the home team has a high points tally prior to the game the number of cards shown during the game is lower, perhaps suggesting that successful home teams need not resort to foul play in order to win. In contrast, a successful away team with a

Table 6
OLS and referee fixed effects models for the number of yellow and red cards per game.

Variables	OLS	OLS	FE	FE
TIER 1	0.667*** (9.85)	0.484*** (6.16)	0.375*** (4.62)	0.230** (2.29)
POST00/01	0.345*** (3.43)	0.268*** (2.64)	−0.012 (0.10)	−0.142 (1.18)
TIER 1 × POST00/01	−0.594*** (7.16)	−0.553*** (6.64)	−0.524*** (5.32)	−0.469*** (4.75)
LOGATTEND		0.219*** (4.49)		0.225*** (4.70)
SCORE		−0.199 (1.56)		−0.015 (1.24)
SCORESQ		−0.028*** (6.55)		−0.027*** (6.47)
DERBY		0.737*** (7.86)		0.739*** (8.05)
MONTHSTOGO		0.073*** (10.09)		0.073*** (10.35)
HOMEPOINTS		−0.099** (2.32)		−0.090** (2.15)
AWAYPOINTS		0.160*** (4.06)		0.176*** (4.58)
CONSTANT	2.753*** (37.89)	0.333 (0.74)	3.203*** (37.74)	0.588 (1.32)
Year dummies	YES	YES	YES	YES
Observations	11,169	10,904	11,169	10,904
Adjusted R-squared	0.01	0.04	0.07	0.10

***p<0.01, **p<0.05, *p<0.1.

Absolute t statistics in parentheses.

high points tally prior to the game is associated with more yellow cards being shown as one might expect if successful teams resort to tougher tactics away from home, or else elicit a more robust competitive spirit from home teams leading to more foul play.

The third and fourth columns of Table 5 are identical to the first and second columns but the models now include referee fixed effects which account for fixed unobservable characteristics of referees, such as their innate talent. These models account for more of the variance in card giving than the OLS models. If the fixed talent of those who become salaried referees was driving the effects of salaried contracts we would expect to see a substantial reduction in the size of the coefficient on the interaction term. In fact, although it falls a little it remains large and statistically significant indicating that the non-random allocation of salary contracts to more able referees is not driving the results.¹⁷ As we noted earlier, some tier 1 matches played after the introduction of salaried contracts were not officiated by salaried referees. The exclusion of these games from the estimation makes very little difference to the results. Similarly, the exclusion of the tier 2 games that were refereed by salaried referees does not affect the results.¹⁸

Table 6 performs the same analyses but this time the estimates relate to all cards given during games, including second yellow cards and red cards. The pattern of the results is very similar. Indeed, the impact of the introduction of salaried contracts appears a little stronger. Once again, the results are not sensitive to the exclusion of games where a tier 1 game was officiated by a non-salaried referee after 2000/01 or those tier 2 games which were officiated by salaried referees.

Table 7 replicates the analysis presented in Table 6, but the sample of matches is confined to those officiated by referees who are

¹⁷ Our identification strategy assumes that the growth in the performance gap we attribute to the introduction of salaries is driven by an improvement in the performance of tier 1 referees, rather than a decline in the performance of tier 2 referees. Following a suggestion by an anonymous referee we ran before–after estimates on tier 1 referees only. These models, which are available from the authors on request, confirm a strong statistically significant decline in card giving by tier 1 referees post the introduction of salaries having controlled for referee fixed effects and other controls already mentioned.

¹⁸ These sensitivity tests are available from the authors on request.

Table 7

OLS and referee fixed effects models for the number of yellow and red cards per game, confined to referees observed in the pre- and post-salary contract eras.

Variables	OLS	OLS	FE	FE
TIER 1	0.551*** (7.30)	0.375*** (4.23)	0.365*** (3.84)	0.236** (2.28)
POST00/01	0.135 (1.01)	0.076 (0.57)	−0.023 (0.17)	−0.104 (0.79)
TIER 1 × POST00/01	−0.462*** (4.94)	−0.395*** (4.22)	−0.522*** (5.19)	−0.461*** (4.59)
LOGATTEND		0.196 (3.45)		0.209*** (3.76)
SCORE		−0.027* (1.86)		−0.023 (1.64)
SCORESQ		−0.028*** (5.74)		−0.027*** (5.77)
DERBY		0.831*** (8.18)		0.823*** (8.28)
MONTHSTOGO		0.080*** (9.79)		0.080*** (10.06)
HOMEPOINTS		−0.050 (1.06)		−0.054 (1.16)
AWAYPOINTS		0.173*** (3.92)		0.191*** (4.44)
CONSTANT	2.893*** (33.72)	0.566 (1.08)	3.148*** (35.23)	0.660 (1.29)
Year dummies	YES	YES	YES	YES
Observations	8516	8301	8516	8301
Adjusted R-squared	0.01	0.04	0.07	0.10

Absolute t statistics in parentheses.

*** p < 0.01.

** p < 0.05.

* p < 0.1.

observed both before and after the introduction of salaried contracts. This accounts for the reduction in the sample size. The OLS coefficients are around three-quarters the size of those presented in Table 6, suggesting that worker sorting accounts for at least some of the performance impact associated with the introduction of salaried contracts. However, the coefficients from the referee fixed effects models in columns 3 and 4 are virtually identical to those presented in Table 6. Taken together these results suggest that, while some of the impact of salaried contracts on referee performance stems from the non-random assignment of contracts to the more talented referees, movement of referees into and out of the profession appears to have had little additional impact on referee performance.¹⁹

The size of the performance gap between referees on salaried contracts and those on match fees is potentially informative with respect to the different mechanisms by which the salary effect occurs. We therefore reran estimates similar to those presented in Table 6 replacing the interaction between tier 1 and the post-2000/01 dummy with an interaction between tier 1 and the years since 2000/01 (running from zero in 2000/01 to 8 in 2008/09). The interactions are jointly statistically significant. The coefficients grow in size over the first few years of salaried contracts but then stabilise until the last two years we observe where the interaction effects fall in size and become statistically non-significant.²⁰ Clearly, the salaried contracts had an immediate and substantial impact but this was not sustained throughout the period. There are a number of possible ways to interpret this result. First, initial competition for salaried contracts was akin to a 'competitive shock' the effects of which were felt early on but dissipated over time once some referees proved successful in obtaining those job slots. Second, and alternatively, non-salaried referees may have begun to improve their relative performance in the

hope of becoming salaried referees.²¹ Either way, it appears that any impact associated with the new training regime has not been sustained over time. Nor have the increased real income returns to salaried contracts incentivized salaried referees to exert still more effort relative to their piece rate counterparts in order to retain those contracts.

8. Conclusions

Using data on referees in the top two divisions of English football we show that the introduction of salaried contracts for those in the top division resulted in improved referee performance, as measured by the number of cards referees give each game. Our difference-in-difference estimator isolates the effect of the switch to salary contracts having accounted for trends common to salaried and non-salaried referees which may have influenced their performance. We account for fixed unobservable differences across referees, such as natural talent, using referee fixed effects. The fixed effects estimates show that some of the effect is due to the non-random allocation of salaried contracts to more talented referees. However, although the effect attributable to salaried contracts falls a little it remains strong and statistically significant. Furthermore, worker sorting in and out of refereeing in tiers 1 and 2 over the period makes little difference to the impact of salaries on performance.

Our result indicates that worker performance improves when workers are paid a salary as opposed to a short-term contract. This is because, in the case of football referees, the introduction of salaried contracts was part of the 'professionalization' of the occupation. This has two clear performance-enhancing effects. The first is a worker sorting effect, with more ambitious and, potentially, more talented referees, seeking to enter and remain in the profession. This accounted for some of the effect we find. The second effect is the 'career concerns' effect associated with the desire to perform well today in the hope of securing or retaining a salaried position in the future. This second effect is closely associated with the much higher income one can command as a salaried referee compared to a referee operating on a match-based piece rate. The moment salaries were introduced in 2001/02 those on salaried contracts were able to earn vastly more through refereeing than would have been the case if they continued to work on a match fee basis. The real income gap grew over time with the appreciation of the value of salaried contracts relative to piece rate payments. Thus the threat of not having one's salary contract renewed was considerable, no doubt leading referees to exert greater efforts to perform well and thus merit contract renewal.

The other reason why one might have seen improved performance among salaried referees relative to those on a match fee is that salary contracts required referees to undergo rigorous off-the-job training to improve their performance. Empirically, it is difficult to distinguish between the career concerns, income effects and training effects on referee performance. However, tests of the differential impact of salary contracts on performance over time indicate that their introduction had an immediate and substantial impact on performance. This effect remained statistically significant until around 2006/07, a finding consistent with a 'one-off' albeit fairly sustained effect associated with professionalization, one which is unlikely to improve further by paying salaried referees more in the future.

Acknowledgements

We thank Michele Pellizzari, seminar participants at Aalto University, Finland, conference participants at the Economics and Psychology of Football conference, Heythrop College and two anonymous referees for the comments.

¹⁹ The same conclusions emerge when we run this analysis for first yellow cards only. Once again, the results are not sensitive to the removal of tier 1 games officiated by a non-salaried referee after 2000/01 nor those tier 2 games which were officiated by salaried referees.

²⁰ These results are available from the authors on request.

²¹ Before–after estimates for tier 2 referees only indicate no clear pattern in card giving in the year dummies, suggesting that tier 2 referees were not improving their performance relative to the pre-period.

Appendix A. Association between ratings of referees and card giving

The Times newspaper publishes match reports every Monday of weekend games played in the Premier League and the Championship. Over the seasons 2001/02, 2002/03 and 2004/04 these reports were accompanied by a grade for the referee's match performance, awarded by the journalist responsible for the match report. These grades vary from 1 to 10 and are normally distributed with an average grade of 5.6. A higher score denotes a better referee performance. The reports typically lack an explanation of the referee's grade but some rationales can be discerned from the match reports. Low marks are awarded where a referee makes a gross error of judgement such as disallowing an obvious goal, or allowing a goal that should not have been awarded. The award of sanctions does feature prominently in several reports. In a game played between Arsenal and Tottenham, November 16, 2002, the referee was given a score of two and was described as 'trigger-happy' in the use of cards, with some of the awards described as 'harsh'. The referee of the game between Nottingham Forest and Newcastle on September 13, 2003 awarded 10 yellow cards and two red cards; his performance was described as 'pompous officiating' and also received a score of two. Similarly, in the game between Nottingham Forest and Reading on December 21, 2002, the referee awarded five yellow cards in a match that was described as "too one-sided to be competitive... and without a single malicious tackle. The referee has succeeded in ruining a bad game." The referee was given a score of one. At the top end of the distribution the performance of the referee in the game between Gillingham and Portsmouth on September 7, 2002 received a score of nine. The journalist commented on "a fully committed but bookings-free match that was superbly overseen by Steve Baines".

Referee fixed effects models indicate a strong negative correlation between referees' grades and card giving. These models, reported below, control for referee fixed effects, tier and season.

Appendix Table
Referee fixed effects models for referee ratings.

Variables	1st yellow	2nd yellow/reds	All cards
1st yellow	-0.170*** (8.70)	-0.158*** (8.04)	
2nd yellow and reds		-0.318*** (3.92)	
All cards			-0.171*** (9.40)
Constant	6.346*** (49.56)	6.385*** (50.00)	6.395*** (50.07)
Observations	1305	1305	1305
Adjusted R-squared	0.10	0.11	0.11

Include fixed effects for referees and dummies for tier and season.

***p<0.01, **p<0.05, *p<0.1.

Absolute t statistics in parentheses.

References

Bandiera, O., Barankay, I., Rasul, I., 2005. Social preferences and the response to incentives: evidence from personnel data. *Quarterly Journal of Economics* 120, 917–962.

Bandiera, O., Barankay, I., Rasul, I., 2007. Incentives for managers and inequality among workers: evidence from a firm level experiment. *Quarterly Journal of Economics* 122, 729–773.

Buraimo, B., Forrest, D., Simmons, R., 2010. The 12th man?: refereeing bias in English and German soccer. *Journal of the Royal Statistical Society, Series A* 173, 431–449.

Corts, K., 2007. Teams versus individual accountability: solving multitask problems through job design. *The Rand Journal of Economics* 38, 467–479.

Dawson, P., Dobson, S., Goddard, J., Wilson, J., 2007. Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanctions in the English Premier League. *Journal of the Royal Statistical Society, Series A* 170, 231–250.

Dobson, S., Goddard, J., 2001. *The Economics of Football*. Cambridge University Press, Cambridge.

Dohmen, T., 2008. The influence of social forces: evidence from the behaviour of football referees. *Economic Inquiry* 46, 411–424.

Fernie, S., Metcalf, D., 1999. It's not what you pay, it's the way that you pay it and that's what gets results: jockeys' pay and performance. *Labour* 13, 385–411.

Forrest, D., McHale, I., McAuley, K., 2008. "Say it ain't so": betting-related malpractice in sport. *International Journal of Sport Finance* 3, 156–166.

Freeman, R.B., Kleiner, M.M., 2005. The last American shoe manufacturers: decreasing productivity and increasing profits in the shift from piece rates to continuous flow production. *Industrial Relations* 44, 307–330.

Frick, B., Gürtler, O., Prinz, J., 2008. Men in black: monitoring and performance of German soccer referees. In: Dietl, H., Franck, E., Kempf, H. (Eds.), *Soccer: Economics of a Passion*. Hofmann, Schorndorf, pp. 309–322.

Garicano, L., Palacios-Huerta, I., Prendergast, C., 2005. Favoritism under social pressure. *The Review of Economics and Statistics* 87, 208–216.

Holmstrom, B., Milgrom, P., 1991. Multitask principal-agent analyses: incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization* 7, 24–52.

James, S., 2008. www.guardian.co.uk/football/2008/aug/14/premierleague12008.

Lazear, E., 1986. Salaries and piece rates. *Journal of Business* 59, 405–431.

Lazear, E., 2000. Performance pay and productivity. *The American Economic Review* 90, 1346–1361.

Lazear, E., 2006. Speeding, terrorism and teaching to the test. *Quarterly Journal of Economics* 121, 1029–1061.

Lemieux, T., MacLeod, W.B., Parent, D., 2009. Performance pay and wage inequality. *Quarterly Journal of Economics* 124, 1–49.

Mukherjee, A., Vasconcelos, L., 2010. Optimal disclosure policy when firms offer implicit contracts. *The RAND Journal of Economics* 41, 549–573.

Pendleton, A., Whitfield, K., Bryson, A., 2009. The changing use of contingent pay in the modern British workplace. In: Brown, W., Bryson, A., Forth, J., Whitfield, K. (Eds.), *The Evolution of the Modern Workplace*. Cambridge University Press, pp. 256–284.

Prendergast, C., 1999. The provision of incentives in firms. *Journal of Economic Literature* 37, 7–63.

Rickman, N., Witt, R., 2008. Favoritism and financial incentives: a natural experiment. *Economica* 75, 296–309.

Sawdon-Smith, R., 2001. Refereeing revolution starts now. www.readingrefs.org.uk/ftmpages/FTM80.html2001.

Shearer, B., 2004. Piece rates, fixed wages and incentive effects: evidence from a field experiment. *Review of Economic Studies* 71, 514–534.

Weston, M., Castigna, C., Impellizzeri, F., Rampini, E., Breivik, S., 2010. Ageing and physical match performance in English Premier League soccer referees. *Journal of Science and Medicine in Sport* 13, 96–100.