

AN EVALUATION OF SPORTRADAR'S FRAUD DETECTION SYSTEM

by

David Forrest

Professor of Economics, University of Liverpool, UK

and

Ian G. McHale

Professor of Sports Analytics, University of Salford, UK

September, 2015

CONTENTS

SUMMARY	1
1 INTRODUCTION	2
1.1 The purpose and scope of this Report.....	2
1.2 Conceptual framework.....	3
1.3 Criteria for assessing screening systems.....	6
1.4 Structure of the Report.....	9
2 THE QUALITY, SCOPE AND RELIABILITY OF THE DATA USED IN THE FDS ..	12
2.1 Introduction.....	12
2.2 Odds Data.....	12
<i>The scope of the data.....</i>	<i>12</i>
<i>The quality of the data.....</i>	<i>15</i>
2.3 Sports Data.....	16
2.4 Synchronisation of Odds and Sports Data	17
3 THE MATHEMATICAL MODELS OF THE FDS.....	19
3.1 The role of the mathematical models.....	19
3.2 The pre-match model	20
3.3 The in-play model.....	23
<i>The in-play model in practice.....</i>	<i>24</i>
<i>Testing the efficacy of the in-play model.....</i>	<i>27</i>
4 THE ALERT PROCESS: STAGE 1 IN THE FDS	29
4.1 The triggering of alerts pre-match	29
<i>The general approach</i>	<i>29</i>
<i>Criteria for creation of an alert</i>	<i>30</i>
<i>Additional criteria for alerts</i>	<i>32</i>
<i>Setting thresholds for alerts ('configuration')</i>	<i>33</i>
4.2 The triggering of alerts in-play	35
4.3 Are the thresholds set appropriately?.....	38
5 STAGE 2 IN THE FDS: HOTLISTING AND ESCALATION.....	40

5.1	Introduction.....	40
5.2	Hotlisting	41
	<i>Description of the process.....</i>	<i>41</i>
	<i>Are the analysts adequately provided with information?.....</i>	<i>42</i>
	<i>Are the analysts appropriately qualified to make the judgement?.....</i>	<i>44</i>
5.3	The escalation process	45
	<i>The input from freelancers</i>	<i>47</i>
	<i>The decision meeting.....</i>	<i>50</i>
	<i>The report on a match classified as suspicious.....</i>	<i>51</i>
	<i>Conclusion.....</i>	<i>53</i>
6	CASE STUDIES: THE FDS IN ACTION	55
6.1	Introduction.....	55
6.2	Case studies.....	56
	<i>Australia</i>	<i>56</i>
	<i>Austria</i>	<i>58</i>
	<i>Estonia and Latvia</i>	<i>59</i>
7	SOME REFLECTIONS.....	62
8	CONCLUSIONS	65
Appendix A:	List of bookmakers monitored within the FDS.....	67
Appendix B:	Odds database and screenshot details.....	71
Appendix C:	Empirically testing the in-play model.	77
Appendix D:	Procedures for hotlisting and escalation.....	81

SUMMARY

In this Report, we examine the likely efficacy of Sportradar's Fraud Detection System (FDS), which monitors betting markets for indications that a (football) match may have been manipulated. As with any screening system, its efficacy is appropriately evaluated by reference to its sensitivity (what proportion of manipulated matches does it identify?) and its specificity (what proportion of matches identified as likely to have been manipulated are true cases of manipulation?).

Our evaluation was based principally on analysing the reliability in construction and execution of each component of the system, both those based on statistical algorithms and those where expert analysts form a final judgement. We also considered case studies of matches known, from external evidence, to have been manipulated

Our conclusion is, first, that, while the level of sensitivity cannot be evaluated precisely, the FDS is likely to identify correctly a significant proportion of manipulated matches. Second, given the quite high proportion of matches identified as potentially suspicious by the statistical algorithms operated by the FDS and the relatively low number of reports finally issued to client organisations, we view analysts at the FDS as being both thorough and cautious when the final decision is taken on whether to classify a match as suspicious. Therefore we believe that the specificity of the screen is likely to be high and that few false positives will be presented.

1 INTRODUCTION

1.1 The purpose and scope of this Report

During the last five years in particular, there has been growing awareness of the scale and level of threat to the integrity of organised sport, including from international organised crime. Several significant Reports on the problem of match fixing have been produced, for example by l'Institut de Relations Internationales et Stratégiques¹ and by the Sorbonne University². Increasingly, governments have also recognised the need for protective measures to be developed, as evidenced by the Council of Europe Convention on the Manipulation of Sports Competitions, adopted by the Committee of Ministers in July 2014.

Against this background, Sportradar offers integrity services to sports federations and competitions and state and law enforcement agencies. Its Fraud Detection System (FDS) monitors betting markets for abnormal activity with a view to identifying fixtures where there is evidence suggestive of manipulation of the event. Reports are then issued to the sports federation or other partner as appropriate. Successful detection of match fixing potentially aids directly in addressing the threat to integrity as it may lead to removal of corrupt personnel from the sport. Strengthening the chance of detection may also serve as a deterrent to sports insiders agreeing to participate in corrupt practices to begin with.

The present authors were contracted by Sportradar to provide an independent evaluation of the efficacy and reliability of the FDS. Our investigation took place between March and May, 2015. Methods employed included reviews of multiple internal documents describing procedures in place at Sportradar, practical testing of systems for assembling odds data, theoretical and empirical evaluation of the statistical models embedded within the FDS, live observation of the work of analysts as matches were played and data generated, and attendance at subsequent meetings where Sportradar personnel took decisions as to whether to report a match as suspicious. Though Sportradar operates the FDS for certain other sports, such as handball and cricket, our investigation was limited to its application to football. Sportradar monitors football betting markets on behalf of UEFA and several other soccer federations and competitions around the World.

Naturally, our perspective and approach were influenced by our academic background in economics and statistics. In carrying out the Evaluation, we were able to draw on substantial experience in producing peer-reviewed papers and reports on sports integrity, the efficiency of

¹ IRIS, 'Sports betting and Corruption: How to Preserve the Integrity of Sport', Paris, 2012.

² Université Paris 1 Panthéon-Sorbonne and International Centre for Sport Security, 'Fighting against the Manipulation of Sports Competitions', Paris, 2014

betting markets and the statistical modelling of football matches. In the past, we have also advised various sports federations and public bodies on relevant related issues.

In this introductory section, we discuss the conceptual framework within which we approached the Evaluation. Then we describe how we broke down our commission into steps such that all the building blocks which make up the FDS were covered.

1.2 Conceptual framework

Forensic economics and statistics has been applied to sports data by a number of authors to attempt to gain a general idea of the prevalence of corruption in various sports. For example, Wolfers³ estimated the proportion of college basketball matches which might have been subject to point shaving (a practice of manipulating the margin of victory such that, while a team wins on the court, wagers on its opponent win on the betting market), Duggan and Levitt⁴ revealed sumo wrestlers swapping wins across tournaments depending on when one competitor most needed a win, and Minor and Brown⁵ modelled ‘tanking’ (playing with low effort) in professional tennis. All such papers may be argued to be informative regarding the rough order of magnitude of the prevalence of various corrupt practices. However, analysis of sports data alone is unpromising at the micro level, i.e. for detection of individual cases of corruption, because data are too ‘noisy’. Data can highlight instances where individuals or teams underperform relative to ‘expectations’; but underperformance is common from the inherent uncertainty of sporting contests and treating a surprise result, or any underperformance, as potentially evidence of corruption would produce a very high number of false positives.

For this reason, it appears a sensible course to search, as Sportradar does, for evidence of corruption by monitoring betting markets alongside sports data. Of course, fixes may be instigated with a variety of motives. There may be a ‘sporting’ motivation, for example one team needs to win to avoid relegation and the owners of the two teams involved agree (for money or for future return of favour) that an appropriate result be ‘manufactured’; or the motive for ‘buying’ a fix from players might simply be to enable winning money on the betting market. But, even in the former case, effects from the fix may be evident on betting markets because insiders cannot resist seeking to profit from knowledge that a match will be manipulated. Thus a variety of fixes, not just those initiated by external parties purely for betting gain, might be detected by procedures which include monitoring betting markets on sports events.

³ J. Wolfers, ‘Point shaving: Corruption in NCAA basketball’, *American Economic Review*, 2006, pp. 279-283.

⁴ M. Duggan and S.D. Levitt, ‘Winning isn’t everything: Corruption in sumo wrestling’, *American Economic Review*, 2002, pp. 1594-1605.

⁵ D. Minor and J. Brown, ‘Selecting the best? Spillovers and shadows in elimination tournaments’, *Management Science*, 2014, pp. 3087-3102.

Monitoring the flow of data from betting markets to detect fraud with the aid of statistical models and algorithms falls within the general field of statistics termed *anomaly detection*⁶. In many applications it is appropriate for the search to focus on ‘outliers’, for example, on cases in data from medical screening where blood pressure is exceptionally high given a subject’s age, gender and weight. Such cases may then be selected for more detailed examination. However, the literature notes that, where the reason for identifying anomalies is to detect possible malpractice, it is very common to focus instead on ‘bursts of activity’. For example, algorithms for detecting credit card fraud are constructed to emphasise the significance of instances where there is a sudden increase in the frequency with which a card is used. This does not necessarily indicate abuse but triggers further inquiry because most cards appropriated by criminals will in fact be used frequently, to maximise returns quickly, before the theft of the card is discovered.

In its emphasis on ‘bursts of activity’, the principles underpinning the FDS therefore put it squarely in the mainstream tradition of forensic statistics. Bursts of activity in the context of a betting market may be captured by observing unusual changes in odds. Such changes will often signify unusually heavy flows of money which reflect that certain bettors believe that the previous odds were favourable to them; sometimes this will be because they themselves had arranged for the match to take a certain course. Changes may also be observed where bookmakers come to form an opinion that a match is being subject to manipulation, for example they may then respond by taking the odds into an untypically uncompetitive range. Thus odds changes capture both bettor and bookmaker knowledge and behaviour and algorithms to identify anomalous odds changes will therefore present a selection of cases where the behaviour of bettors and bookmakers indicates a need for further investigation.

Perhaps the closest analogue to the FDS lies in the activity of agencies charged with detecting insider trading on stock markets. Indeed the analogy is almost perfect to the extent that fixers in sport are also seeking to trade on a financial market (betting) to profit from specific private information (for example that players have agreed to concede a certain number of goals). Insider trading watchdogs on stock markets, similar to Sportradar, “primarily look for suspicious trading patterns, usually with a combination of sophisticated software systems, rules of thumb and common sense”⁷. The procedures described below for picking out matches as suspicious indeed embody use of both sophisticated computer algorithms and employment of judgement by experienced specialists.

⁶ A comprehensive survey is provided by V. Chandola, A. Banerjee and V. Kumar, ‘Anomaly detection: A survey’, *ACM Computing Surveys*, 2009, Article 15.

⁷ N. Mehta, ‘The ins and outs of insider trading’, *Financial Times*, July 7, 2013.

But in some sense analysts at Sportradar face a tougher problem than those of such as the US Securities and Exchange Commission (SEC). Typically, national agencies such as the SEC are empowered to obtain information on the identity of traders in cases when anomalies such as heavy buying before a favourable announcement are observed (whereas this is not possible in the betting market, where transactions often take place in regions with no effective regulation). Further, the SEC is permitted itself to seek further evidence using law enforcement techniques such as wire-tapping. Nevertheless, an ‘official’ speech on behalf of the SEC⁸ noted that it was still rare to find a ‘smoking gun’ and so “insider trading is an extraordinarily difficult crime to prove” because the evidence is very commonly just circumstantial. The SEC used American experience to argue the importance of legislating insider trading as a civil as well as a criminal offence, as America has done. Otherwise, deterrence of insider trading will fail because criminal guilt beyond reasonable doubt will too often not be established whereas liability in the civil courts can be adjudicated on the basis of the balance of probability. Thus SEC investigations sometimes lead to criminal but more often to civil penalties. By analogy, screening for match fixing will sometimes lead to criminal prosecution (as we will illustrate below) but it is realistic to expect that offences will more often have to be considered within the framework of disciplinary proceedings within sport.

In any case, analysis requires not just data to identify bursts of activity on the relevant financial market but also detailed information on the real events that prices on financial markets reflect. Just as with sudden price movements on the stock market, *most* sharp changes in odds may be linked to events which make them ‘rational’. For example, pre-match odds on a football match may shift abruptly an hour before kick-off when team line-ups are announced and markets then re-evaluate outcome probabilities because of surprise omissions from one of the teams. Therefore, to distinguish between ‘rational’ and ‘perverse’ or ‘suspicious’ price movements, analysts need access to reliable sports data as well as sight of trends in the betting market. It is also important that there is consistency in the time-stamping of sports and betting data. For example, in the in-play market odds will normally shift substantially as soon as there is a goal given that soccer is a low scoring sport and one goal is therefore typically significant for the probabilities of final outcomes. A ‘normal’ odds shift following a goal would not therefore be suspicious whereas the same odds movement shortly before the goal may justify closer scrutiny. For all these reasons, our review had to pay great attention to the quality of data flowing within the FDS and to the consistency of the time-stamping of these data. Again these are precisely the issues faced in the monitoring of other financial markets by such as the SEC.

⁸ Speech by T.C. Newkirk at the 16th International Symposium on Economic Crime, Jesus College, Cambridge, retrieved from www.sec.gov

1.3 Criteria for assessing screening systems

The FDS is a screen for identifying fraud. It tests whole populations (for example, all matches in a given competition) and its output is essentially to declare which matches have tested positive and which have tested negative. In the general literature on screening, it is conventional to judge the usefulness of any screen test against the criteria of *sensitivity* and *specificity*. These criteria relate to the proportions in screen results of true positives, false positives, true negatives and false negatives. Here:

A true positive would be a match which the FDS labelled as suspicious and which had indeed been manipulated.

A false positive would be a match which the FDS labelled as suspicious but which had not in fact been manipulated.

A true negative would be a match which the FDS had not labelled as suspicious and where indeed nothing untoward had taken place.

A false negative would be a match which the FDS had failed to identify as suspicious but where fixing had in fact occurred.

A screen is said to have high *sensitivity* when it classifies as positive a high proportion of cases where the condition of interest is present, i.e. a sensitive screen does not miss out many true cases.

A screen is said to have high *specificity* when it correctly classifies as negative a high proportion of cases where the condition of interest is absent. High specificity indicates a low probability that a positive test result is incorrect.

Generally, there is a trade-off between sensitivity and specificity. Where there is a higher *cut-off* (or threshold) in the specification of criteria which determine whether the screen declares a positive, this has the cost of raising the proportion of true cases missed by the exercise (sensitivity is weakened). But a higher cut-off will normally improve specificity in that fewer false cases will then be included in the set of cases subject to further investigation. In the general case, the choice of cut-off will determine the relative degrees of sensitivity and specificity attached to the screen; and in specifying the cut-off users will need to weigh the relative costs associated with false negatives and with false positives. For example, in a medical application of detecting a particular condition, it might be decided that missing cases (false negatives) was not very costly because doctors could not anyway offer effective treatment for the particular condition whereas false positives were costly because follow-up procedures to determine whether a case was a true positive were invasive and traumatic for patients. In this sort of

circumstance, the cut-off would be set very high or it might even be decided that the screen should not be used at all.

But, where sensitivity and specificity are both judged important, a recommended approach to avoid harm from trading-off between them is to introduce a two-stage screening⁹. The first-stage screen is constructed to exhibit high sensitivity (but consequently low specificity). This makes it likely that few cases in the population are missed. Cases testing positive at this stage are then subject to a second-stage screen designed with the emphasis on specificity. The intention is to eliminate a high proportion of the false positives generated by the first-stage screen. The combined result from the two screens should then satisfy both desirable criteria, sensitivity and specificity.

In some applications, the sensitivity and specificity of a one- or two-stage screen can be evaluated numerically. Sensitivity and specificity are then typically measured as:

Sensitivity= (# of true positives) / (# of true positives + # of false negatives)
= probability that a true case is classified as true

Specificity= (#of true negatives)/ (# of true negatives + # of false positives)
= probability that a false case is classified as false.

In some circumstances sufficient information emerges after testing for a precise numerical evaluation of these probabilities to be made on the basis of historic data. For example, algorithms to detect credit card abuse will fail to identify some cases where a credit card has been stolen and is currently being used by a criminal. But almost all cases of missing credit cards will eventually be noticed and reported by the customer and so the number of false negatives over a sample period will become known *ex post*. Sensitivity and specificity can therefore then be precisely measured to provide yardsticks by which the utility and efficacy of the screening procedures may be judged.

In the present application, to the FDS, such precise numerical evaluation is not possible. In fact, sensitivity appears essentially unknowable in this case since any match fixing not revealed by the screen is unlikely to be revealed subsequently. In principle, of course, there may in future emerge

⁹ See, for example, A.G. Laikhen and A. McCluskey, 'Clinical tests: sensitivity and specificity', *Continuing Education in Anaesthesia, Critical Care and Pain*, 2008, pp. 221-223.

a set of matches uncovered independently of the screen, for example if police stumble across a criminal organisation with records of which matches it had fixed. It could then be ascertained whether those matches had tested positive or negative in the FDS. This would be a fair test of the sensitivity of the FDS always providing that the sample size was adequately large. But to date there do not exist sufficient numbers of independently discovered known proven matches from the limited period for which the FDS has operated for any measurement to be treated as a wholly statistically valid estimate.

Nevertheless some indicative evidence was available to us in the form of case studies where law enforcement had been involved in investigation of match fixing. There were some cases where police had requested information from Sportradar regarding matches where fixing had been verified in independently initiated investigations and to the point of criminal proceedings being considered. If the FDS had in fact flagged up such matches at the time they were played, this would be suggestive of ‘good’ sensitivity even if the sample size was too small to make a serious numerical estimate of the sensitivity index.

Similarly, regarding specificity, relatively few reports from Sportradar lead to prosecution but this does not imply that they were not true cases. Sports federations or law enforcement, for their own reasons, may not take any follow-up action. For example, the federation may be nervous of reputational damage to the sport or police and prosecutors may operate in a jurisdiction where there is no clear offence with which to bring charges. This is an obstacle to calculation of a specificity index. But, if a national federation does pass on reports to the police and these do result in verification of malpractice, this would be at least suggestive of ‘high’ specificity.

Even where empirical evaluation of sensitivity and specificity is not feasible, the concepts should not simply be ignored. They are the essential conceptual criteria by which the efficacy of any screen should be judged. Our strategy for assessing sensitivity and specificity was to make use of case studies to provide indicative evidence but also to investigate closely issues such as the coverage of the data and the choices of cut-off points in the system’s specification. For example, it is intuitive that the FDS would be insensitive (i.e. would often miss cases of fraud) if it monitored only a narrow range of betting platforms or if it monitored only well-regulated betting environments where criminals would be unlikely to place bets. Similarly, classification of matches as suspicious would be unlikely to be adequately reliable if the data employed in the testing procedure contained significant errors. Therefore examining the set-up of the FDS in terms of detailed specifications and reliability of inputs also makes it feasible to form an informed view of how confident clients can be in the sensitivity and specificity of FDS systems.

In general, we shall report below that the FDS breaks down into a number of consecutive stages and that parts of the system emphasise sensitivity and parts emphasise specificity. This is

consistent with the preference derived from the medical literature for the use of two-stage testing procedures as noted above.

1.4 Structure of the Report

Following an initial inspection of the FDS internal documentation, we determined that Sportradar had implicitly adopted an appropriate two-stage procedure, leading to final decisions on classification of some matches as positive (i.e. suspicious). In the first stage which is automated, screening yields *alerts* for a relatively high proportion of matches, which makes it plausible that sensitivity will be high. In the second stage, which is broken down into two parts (*hotlisting* and *escalation*), judgemental evaluation is applied in a systematic way to discard evident false positives with the goal of assuring high specificity when cases are finally classified as positives (suspicious matches).

1. Throughout the betting period for each match, betting odds are obtained at high frequency from many platforms. Algorithms developed by Sportradar trigger e-mail ‘alerts’ to be considered by one of the duty analysts in the London (or Hong Kong or Sydney) office. These alerts can be considered as the first-stage screen. In the pre-match betting market, alerts are mostly linked to changes in odds above a threshold (or cut-off) specified for the relevant football league¹⁰ (‘odds’ in the FDS algorithms are expressed as a statistic termed ‘netwin’¹¹). For a subset of leagues, alerts may additionally be triggered by deviation (beyond a threshold) of observed odds from ‘true’ odds where true odds are generated from a Sportradar probabilistic forecasting model based on Elo ratings of teams. In the in-play market, alerts are linked to significant deviation of odds from those predicted (‘calculated probabilities’) by a statistical forecasting model where outcome probabilities depend on pre-match odds, the current score, the time remaining and whether teams are currently short-handed because of red cards. Additional criteria for alerts to occur include withdrawals of market by bookmakers and unexpectedly high volume on Betfair.

To test the reliability of these components of the FDS required us to consider whether the coverage of the betting market was adequate; to verify the accuracy of the processes for collecting data from both the betting market and the sports event; to check for appropriate synchronisation of the timing of data on odds and scoreline; to examine back-up procedures to be used in the event of failure of the automated procedures for collecting data; to check the validity of defining odds by the statistic ‘netwin’; to assess whether the thresholds set for alerts

¹⁰ Thresholds are higher in low-status leagues where the liquidity in the associated betting market is typically lower and thus even relatively small money flows may shift odds markedly.

¹¹ *Netwin* is the profit that would accrue to the successful bettor for a one unit stake. To illustrate, suppose the bookmaker quoted ‘decimal odds’ (now the conventional way of quoting odds) of 1.5. This means that a successful bettor would have a claim to 1.5 units of money per unit staked. But part of this is return of stake: the profit or ‘netwin’ is only 0.5 money units per unit staked. Thus netwin= odd-1. In fact, it is also the number that used traditionally to be quoted by British bookmakers who would express the odds as 1/2 (=0.5).

are sensible; to consider whether the specification of the statistical models conform with best practice in the field; and to test the empirical performance of the statistical models.

2. Where an alert or multiple alerts on the same event are generated, the analyst sees the ‘deviation’ in odds in tabular and/or graphical form (also information on whether any bookmakers have withdrawn coverage of a match). At this stage, the system gives him access to any additional information available, for example on latest sporting data, such as team news, and any relevant recent reports from correspondents (‘freelancers’) engaged to report significant football stories to Sportradar on a routine basis. The analyst may also use online sources to research possible factors triggering the alert. Then the analyst must use skill, knowledge and experience to judge whether the alert can be dismissed or should remain under active consideration in the FDS. Knowledge of both sport and betting has to be used in decision-taking. For example: in the in-play market, a deviation from the calculated odds might be explained by a red card having been awarded to an unusually influential member of the team (the statistical model takes account of the award of a red card but not of the identity of the carded player); and whether bookmakers failing to offer in-play betting on a match is significant depends on what is the usual commercial practice for those bookmakers for that League. The analyst therefore needs to draw on his knowledge of sport and of betting markets. Having reached a judgement based on such considerations, reasons must be logged in the FDS. If the analyst cannot find adequate legitimate reason for the alert and so still finds the match potentially suspicious, then, subject to agreement by a supervisor, the match is ‘hotlisted’ for further consideration in the FDS. This process can be interpreted as the first part of the second-stage screen, filtering out false positives from the first-stage that can be judged as such by appropriately experienced personnel aided by appropriate collated information.

We obtained documentation of the career paths of each analyst to inform our judgement of whether their background and skills equipped them for the task of accurately weeding out false positives. We observed them at work to help us understand this stage in the operational process and form a view on how effectively and reliably it was carried out. A new information source not covered at stage 1 was the system of freelancers employed by Sportradar to cover each country where it has a client competition. We reviewed the procedures used by Sportradar to control the quality and relevance of information flows from freelancers.

3. Details of any match which has been hotlisted are flagged up in the FDS to allow other analysts to consider the case. Meanwhile, the supervisor sends questions seeking further information about the match from the freelancer in the relevant country. The questions broadly follow a set pattern but are modified in each case to be made specific to the match under suspicion. Responses are expected within 24 hours. In some cases, questions are also sent to the ‘scout’ who attended the game (Sportradar has a scout at many matches, with the primary responsibility of supplying reliable sports data, such as time of kick-off and principal match events). Sports data are validated and further analysis of personnel in the match is conducted.

Once all the relevant documentation has been assembled, normally on the day after the match, analysts and supervisor compare the facts against a ‘suspicious betting checklist’. They debate the circumstances of the match and reach a consensus on whether Sportradar should issue a warning to the client sports federation or competition. The number of personnel involved in this debate varies according to duty periods but at least three analysts must formally agree for the process to move towards a warning being issued, which would signify that the evidence indicated strongly that the match had been manipulated. Normally many more than three are involved in decision-taking and sometimes opinions are sought from other offices. In the event that a decision to escalate is taken, the team will allocate one of three warning levels (for use internally) and the analyst responsible for the match will write a full report, which is subject to a checking process. In the alternative case where the judgement reached is that there are insufficiently strong grounds for concluding that the match had been manipulated, the analyst writes a detailed statement to interpret/ explain what was observed in the betting data. This statement is checked by the supervisor and then referred to the weekly ‘Escalation Review’, which has the option to report the match after all.

Our review of this second part of the second-stage screening, termed the ‘escalation process’ by Sportradar, included attendance at the discussions on four hotlisted matches (for two of which it was decided to issue warnings, i.e. two ‘positive’ final test results were declared) as well as interviews with analysts and inspection of internal documents describing the procedures to be followed. We also inspected a sample of questions sent to freelancers and their replies. Other documents subject to review included the manual setting out how a report on a suspicious match should be written and what information it should include. We obtained data on the number of matches which triggered alerts, the number which were hotlisted and the number for which a warning to the football authority was issued. We used all these sources to form a judgement on the efficacy of the procedures followed for deciding on which matches from the previous steps in the review process were finally to be categorised as positives according to the FDS screen.

We have provided here a basic outline of how the FDS works (more detail will be presented below) in order to draw out some of the key tasks which we judged would have to be included in our Evaluation. Subsequent sections of the Report will focus closely on particular components of the overall task.

Thus, Section 2 covers all issues related to the range and quality of data concerning both the betting market and the sport. Section 3 presents a formal review of the statistical models embedded within the FDS. Section 4 examines the alerts process including alert logic and rules. The next stage in the FDS is for an analyst to decide whether a match subject to an alert or alerts merits more detailed consideration: this ‘hotlisting’ stage is described and discussed in Section 5, which also reports on the next stage, the escalation process. Section 6 considers case studies of matches where there is police or judicial evidence of manipulation and how the FDS performed with respect to those matches. Section 7 offers some reflections on our exercise and Section 8 summarises our conclusion. Appendices present technical details and a diagrammatic representation of FDS processes to assist readers further in understanding how the FDS works.

2 THE QUALITY, SCOPE AND RELIABILITY OF THE DATA USED IN THE FDS

2.1 Introduction

The essence of the FDS is to monitor for irregular (and potentially suspicious) activity in betting markets where what is to be regarded as irregular cannot be defined independently of the situation in the sports event (for example, the strength of the teams for the pre-match market, the latest score for the in-play market). It is therefore a necessary condition for the efficacy of the FDS that appropriate data are collected from both the betting and sports sectors and that there be adequate assurance that all the data input into the FDS are reliable. In this section, we review first the assembly of betting data, then the assembly of sports data and, finally, the important process of ensuring consistency between them in respect of the timing of events occurring in each sector.

The emphasis in this section is on data collection and processing prior to and during the match. These data feed directly into Stage 1 of the FDS where algorithms identify circumstances where what is observed justifies the sports event being moved into the first part (and then possibly the second part) of Stage 2, where it will be given further consideration by analysts as a potentially manipulated event. During the stage 2 procedures, additional information may be sought by analysts to inform their judgement. We will comment on the quality of the supplementary information gathered at those points in later sections. The present section focuses on the data collected before and during the match and fed directly into the system, to be used by the algorithms in automatically creating alerts according to parameters set within the FDS.

2.2 Odds Data

The scope of the data

Steadily, since the Millennium, and in response to development of technology conducive to remote gambling, the sports betting market has evolved into a truly global financial market. For example, betting on any football match in Europe will be offered by bookmakers located all over the World and by many different types of bookmaker, such as European or Asian, licensed or illegal, state-owned or private-sector. In the contemporary globalised world, whichever way the market is segmented, one sector does not operate in isolation from the rest. For example, a surge of money on one side of a bet in Asia will shift odds in Asia. But, following the odds shift triggered by weight of new money in Asia, sophisticated private traders, perhaps using automated trading, may seek to exploit any resulting gap in odds between Asia and Europe (arbitrage); where the odds shift is large, European bookmakers may proactively adjust their own odds to close down arbitrage opportunities and to reduce risk. As a consequence of these activities, *significant* odds movements in Asia will typically be echoed on European markets

very quickly afterwards. The process is the same as in any other financial market where technology permits traders in one region to access markets in another region. If ‘something is happening’, ripples will be observed in many sub-sectors even if the relevant initial trades were all executed in one centre. Consequently, suspect matches in FDS often trigger multiple alerts (across bookmakers) and alerts may also occur in response to flows of tainted money which were *initially* placed with non-observed operators. Complete coverage of *all* bookmakers is not therefore a pre-requisite for the efficacy of the FDS.

In fact, the number of bookmakers operating in the World is unknowable since those located in countries where betting is illegal (for example, China, India, USA) are not registered with any authority. Any monitoring exercise must therefore rely on observing odds movements in only an observable subset of bookmakers which offer wagers on a particular event. Inevitably this runs the risk that some small-scale home-made fixes will escape detection. However, any resulting weakening of the sensitivity of the monitoring is likely in practice to be slight and not to affect the ability of the system to detect significant fraudulent activity. The FDS brings together data from (currently) 286 betting websites (listed in Appendix A) located in Asia, Australasia, the Americas, Africa and Europe. Its coverage in terms of types of operator is equally comprehensive, including state lotteries in Europe which offer sports betting, private sector providers of both land-based and remote betting in Europe, all major trans-national operators, a number of unambiguously illegal operators which are not licensed anywhere, and Betfair¹², by far the dominant ‘betting exchange’. Crucially, coverage also extends to the largest Asian bookmakers including SBOBET and MAXbet.

SBOBET and MAXbet are licensed in Cagayan, Philippines and are the largest bookmakers in the World, each with several times the annual sports betting turnover of major European providers such as William Hill. These are legal operators in the jurisdiction where they are based but are often described as occupying ‘grey markets’ because they draw in money from the illegal markets across Asia: local bookmakers across Asia are unable to bear liability risks associated with large bets and risk is managed by *risk pooling*, which in practice means passing on such bets upwards through a hierarchy of sports books and agents such that, eventually, a large proportion of illegal bets reaches the Cagayan operators. By monitoring SBOBET and MAXbet, and other significant trans-national operators which serve a similar function, the effective coverage of the FDS is therefore extended across a large region where bets are made illegally and where, in fact, a high proportion of World stakes on (say) European and Australian football matches are placed. From the criminal trials of match fixers, we know that betting associated with large-scale fixing is in practice nearly always channelled through Asia because larger bets

¹² An important feature of the Betfair data is that FDS is able to observe not only odds but also volumes transacted. In the case of Betfair, FDS procedures trigger alerts not only when abnormal odds changes are observed but also when the level of activity is unusually large.

are accepted there and lack of meaningful regulation ensures that funds cannot be traced back to source in the event that a match falls under suspicion.¹³

The sensitivity of the FDS, at least with respect to sizeable fraud, appears therefore unlikely to be compromised by insufficient coverage. If large bets are placed with any operators not covered by monitoring, it is probable that these relatively small bookmakers will pass on the bets up the chain to avoid risk and they particularly have an incentive to do so if they suspect they are dealing with fixers (indeed they may add their own funds so as actually to profit from the fix). Thus nefarious money is very likely to enter the observed sector of the Asian market.

The fixers' money is not observed directly but through odds changes. These will occur because the major Asian operators hold to a 'book balancing' business model where they seek to equalise liabilities across sporting outcomes.¹⁴ If there is a surge of money on one outcome, they will reduce exposure to that outcome by adjusting odds and may sometimes also hedge with other operators, causing secondary changes in odds which will again be observed in the FDS.

The set of bookmakers for which odds are monitored by the FDS is therefore large, sufficiently comprehensive to pick up the effects of suspicious money flows in Asia in the markets where they are placed, and comprehensive enough also to detect local fraud in other regions including Europe where unsophisticated offenders may bet with familiar local operators.

Taking all these factors into consideration, we judge that the breadth of coverage of betting platforms is sufficient for us to be confident in its ability to detect a high proportion of significantly-sized fraudulent activity. That not all of the betting market is observed directly

¹³ This is not to say that observation of operators in Europe is redundant even in cases where the primary focus of the fixing operation is in Asia. For example, criminals arrange that players should deliver a certain outcome in an Italian football match and plan to make their illicit profit by wagering on that outcome in Asia. But the players concerned know about the fix and they or their families and associates to whom the information has leaked place local bets to benefit personally from the corruption. The local bookmaker receives heavy betting on one outcome and alters odds as part of risk management. The odds change is picked up by the FDS. Probably this will be the first sign of a fix because the professional criminals will bet late to avoid alerting the market. That the abnormal activity in the Asian market was preceded by unusual localised activity in the country where the match takes place would make the conclusion that a match was fixed more compelling.

¹⁴ Empirical evidence that Asian bookmakers aim to maintain balanced books rather than take positions is provided in A. Grant, J.E.V. Johnson and T. Oikonomidis, T. (2013). Bettors vs. Bookmakers :1-0 ! Examining the origin of information in football betting markets. Working Paper, University of Sydney. This paper also includes convincing econometric evidence that changes in odds observed at SBOBET have strong predictive power in a forecasting model for match results in Europe, indicating that those with valuable, relevant information find it advantageous to wager at providers where the funds eventually arrive in the Phillipines-based bookmakers. Fixers are one class of bettors possessing valuable, relevant information.

appears unlikely seriously to weaken the sensitivity of the FDS (and of course it does not affect specificity at all).

The quality of the data

The FDS uses odds data from both the pre-match and in-play markets. Many bookmakers provide a live feed of odds to Sportradar voluntarily. For these bookmakers, no checking of the accuracy of the odds is necessary – the information comes directly from the source, identically to how it appears to clients using the particular website. On the other hand, some bookmakers, particularly Asian bookmakers, do not provide Sportradar with a live feed of odds. For these other bookmakers, which include some of the World’s biggest operators, Sportradar uses web-crawlers to scrape the data from the bookmaker’s website.

Web-crawlers are automated pieces of software that scrape information off websites. During any match being monitored by the FDS, a web-crawler visits each bookmaker’s website and scrapes the odds every minute, or immediately following a goal or red card event. Web-crawling of websites which change or update information are notoriously difficult to scrape data from since small changes in the positioning of information on a webpage will result in the data being scraped incorrectly. As a consequence of this, information from the FDS crawlers are used in tandem with automated screenshots of the webpages. These are error free in that they are a ‘photograph’ of what the webpage showed at a certain time.

The FDS uses these screenshots as a backup data source in the event that the web-crawlers fail. Here we use them to validate the accuracy of the data collected by web-crawlers. Our experiment is simple: for a sample of matches and leagues and bookmakers, see whether the screenshots agree with the data scraped from the bookmaker’s website.

Screenshots are taken once per minute, whilst odds are recorded at a higher frequency. As such, there are sometimes small discrepancies between the value shown on the screenshot and the value recorded in the scraped data. In cases where the discrepancy lasted for only a few seconds, we treat the two sources as identical.

Appendix B presents a table showing the identity of the matches and the time points at which the screenshot and crawled odds were compared. For each of the 148 matches, data were collected on either the 1X2 market or the Asian Handicap market. For odds collected on both of these markets, three pieces of odds data are recorded. For the 1X2 market the information is the odds available on the home win, draw, and away win match outcomes, whilst for the Asian Handicap market the information includes the over-under prices, and the amount of handicap offered. Out

of 148 individual matches and 444 pieces of information, consistency between the odds on the screenshot and the odds captured by the web crawler was found in all cases.

To get an idea of how likely this is, suppose there is a 1% chance of an error in the webcrawling data. Observing no errors in 444 experiments (or 444 out of 444 correct pieces of information) has a probability of occurring of just 0.0115.

We view this as strong evidence that the web crawlers have been deployed in the FDS to perform their functions correctly, for example collecting the ‘right’ odds on the webpage. We can be confident that the automated capture of odds data using web crawlers works virtually as well as if the relevant operator were supplying odds from its website directly to the FDS.

We also satisfied ourselves that Sportradar has appropriate software in place to detect any failure of equipment which might occur and appropriate written procedures in place to ensure timely remedial action. The relevant internal documents are covered by ISO9001 certification.¹⁵

2.3 Sports Data

In the FDS, sports data relate to information about events occurring which are solely related to the teams and players of the football leagues being monitored. Match dates, teams involved, final scores, timings of goals, team line-ups, details of red and yellow cards (player names and their timings) are all examples of sports data.

As noted already, accurate sports data are a necessary component of an efficacious system of monitoring betting markets. Irregular betting activity cannot be identified as such unless it is compared with what the activity would be expected to look like given the sporting situation.

For many matches, the primary source of live, real time sports data fed into the FDS is the network of scouts employed by Sportradar. We were advised that, of matches subject to FDS monitoring in 2015, scouts watched approximately 39% at the ground. The role of scouts is

¹⁵ ISO9001 certification is awarded by authorised independent auditors who have satisfied themselves that quality management procedures in the company or institution inspected comply with best practice standards as set out in an international agreement concluded in 2008. Sportradar’s certification was awarded in July, 2014 with registration number TIC 15 100 148923. There is annual auditing and certification has to be renewed every three years to ensure continued compliance with quality management standards.

directly to input major match events such as the occurrence of goals and red cards as they occur. The information they present is therefore immediate. FDS also collects live data for all matches (scouted and non-scouted) from bookmaker and other external websites by means of livescore crawlers .

Naturally, there is sometimes a discrepancy between the information for the same event provided by different sources. This could arise because a website has an error (for example, showing 10-0 instead of 1-0 as the score). More routinely, discrepancies between sources will be present when one has already registered a new incident such as a goal whereas others have not yet caught up. For example, one source may display the score as 2-0 when others still have 1-0.

An automated system is in place to select the most reliable current information from the array of sources, scouts and external, used by the system. Very detailed algorithms apply principles such as prioritisation of some sources over others on account of a record of greater reliability, filtering out of information with obvious errors, such as a change in score from 3-1 to 3-5 in one step, and treatment of the first source reporting a new goal in the match as the best source until there is reason to suppose otherwise.

While Sportradar has invested heavily to ensure the accuracy of data driving the automated part of the FDS, It may be noted that data are further checked across sources, and additional information requested from freelancers, for any match passed on for further consideration as potentially manipulated. Therefore no decision on whether a match should finally be classified as suspicious by the FDS can be made before post-match verification of the sports data for that match. Post-match checking of data and the quality of additional information gathered at this point will be considered further in the analysis of post-match procedures (Section 5.3 below).

2.4 Synchronisation of Odds and Sports Data

A difficult task in the FDS is that of synchronisation of the odds and sports data. This is essential to determine whether odds changes occur at an appropriate time, for example just after a goal. This requires rather precise conformity (with respect to the timing of events) between betting and sporting data fed into the system.

To synchronise odds and sports data, Sportradar first identifies the ‘time-stamp’ for the start of the match. To do this relies on some bookmaker’s websites giving a ‘match-clock’ as well as live odds. Sportradar then backwards calculates the bookmaker’s game start time by taking the

timestamp of when the match-clock was crawled, minus its value at that time. This results in one “match start time” timestamp for each bookmaker. The median of all these timestamps is used as the game start time in the FDS.

As a match progresses and odds change, the FDS must know exactly when the odds were available to bettors. For odds that are crawled, the information includes a timestamp of when they were crawled, i.e. when those odds were available for betting. To synchronise the odds data and the sports data, the FDS creates a timeline as described in the following example. Suppose a match is scheduled to start is 18:30:00. The scout at the venue reports the match actually started at 18:31:46. Next, suppose odds were crawled from a bookmaker’s website at 18:43:38. The match clock does not stop in football and so it is known that 12m00s into the match will be at 18:43:46. This means the odds scraped at 18:43:38 were offered in the 12th minute of the match.

The care taken in the recording of timing appeared to us to be sufficient for purpose. This impression was confirmed when we examined the evolution of odds observed during a number of football matches and compared them with the paths of probabilities indicated by a statistical forecasting model applied in-play. When the comparison of responses in each series is made on charts with (FDS) time on the horizontal axis, shifts in odds/ outcome probabilities coincide, suggesting effective synchronisation of information from the sports field and the betting market. This will be illustrated in Section 3 below.

3 THE MATHEMATICAL MODELS OF THE FDS

3.1 The role of the mathematical models

As noted above, the first stage of the FDS is entirely automated and its purpose is to identify cases (alerts) where it will be justified and worthwhile for analysts to examine further the circumstances and betting patterns surrounding the match. This ‘alerts’ stage, to be examined in detail in Section 4 below, is driven by algorithms and underlying the algorithms are mathematical/ statistical models which are used to separate out the abnormal from the normal. The efficacy of the FDS thus requires that the quality and performance of the models embedded in its system makes them fully fit for purpose. The preceding section concluded that the assembly of the data to be used in the alerts process and inputted into the models was perfectly satisfactory; but the validity of the whole exercise demands also that the data then be processed within appropriately constructed models.

Two models are examined in this section.

The FDS tests for irregularities in both the pre-match and in-play betting markets. In the pre-match market, primary reliance is placed on identifying large movements in odds; but, for some of the leagues, the procedures also pick out cases where there is significant divergence between the odds observed in the market and what the odds “should be” according to a probabilistic forecasting model of match outcomes (adjusted to reflect bookmaker vigorish or over-round). It is this probabilistic forecasting model, the *pre-match model*, that we examine first (section 3.2).

More important still, because the bulk of betting volume occurs during a match and criminals will often be wary of placing bets in advance in case it alerts other traders to their nefarious activities, is the *in-play model*. This tracks the expected evolution of odds during a match as events such as goals and red cards occur. Alerts are triggered when there is significant deviation between odds on an individual betting platform and the odds (termed *calculated odds* by Sportradar) predicted by the model. Section 3.3 reports on the in-play model.

For each model, we consider whether the principles employed in its construction accord with best practice in the relevant (and considerable) academic literature in sport analytics. In addition we conducted an empirical test on the performance in practice of the in-play model.

3.2 The pre-match model

The three most important markets monitored by the FDS are: the 1X2 (home win/ draw/ away win) market, the total goals (or over-under) market and the Asian Handicap¹⁶ market. To derive estimated market odds for these markets before the match has taken place, Sportradar employs a statistical model. This ‘pre-match model’ uses information on the two teams’ past results to estimate for each team the probabilities that it will score 0 goals, 1 goal, 2 goals, etc. From these probabilities, expected odds in each of the three markets can then be derived.

A statistician choosing a statistical model is very much like a joiner selecting a tool from his toolbox for a particular task – there is usually a tool that is ‘just right’ for the task in hand. If two pieces of wood need to be joined by a nail, the joiner reaches for his hammer; if a statistician needs to model the number of goals in a football match, he reaches for the Poisson distribution, or a close relation.

The Poisson distribution describes the probability of a team scoring a certain number of goals in a match. This probability is governed by a ‘rate’ parameter, λ , which represents the rate at which goals are scored by a team in a match. As λ increases there is a higher probability of more goals being scored by the team. Of course, the rate of scoring for a team in a specific match depends on many factors: the relative strengths of the two teams, home advantage, form, and the weather conditions (e.g. windy, rainy, hot, cold) to name but a few. It makes intuitive sense that stronger teams will have higher rates of scoring than weaker teams. The task faced by the statisticians at Sportradar is to estimate appropriate values of the scoring rates for each team in a match.

Estimating goal scoring rate parameters for teams has been the subject of academic interest for many years. The basic methodology is to take a weighted average of past goals scored by a team, accounting for which teams they had played. The model can account for (i) *form* of the team in that more recent results influence the estimated scoring rate more than matches further in the past, and (ii) *strength of the opposition* in that goals scored against top teams, as opposed to goals scored against bottom teams, result in higher estimated rate parameters.

¹⁶ In the Asian handicap market, outcomes are binary. Asian Handicap (AHC) is a form of football betting that uses a handicap to approximately equalise the probability of a bet on either team being successful. AHCs use either a whole goal, a half or a quarter of a goal which acts as an advantage given to one of the teams. Bettors then place a bet on the outcome of the match, bearing in mind the handicaps. E.g. Team A +0.5 means that it receives half a goal advantage – so a draw would mean that a bet on A would in fact be a winning bet. Conversely Team B might ‘give away’ a handicap such as -0.5 meaning it must win by at least a goal for betting on that team to be successful. Handicaps can be large if one team is considerably stronger than the other. If Team B’s handicap is -2.5, it must win the match by at least three goals for betting on that team to be successful.

Regarding the efficacy of the model of choice for statisticians modelling football scores, the Poisson distribution, many authors have suggested modifications to account for the particularities of football. For example, Maher¹⁷ disaggregates the scoring process in a football match to include attacking and defending abilities of the two teams playing. Dixon and Coles¹⁸ adapt Maher's model to account for dependence between the goals scored by the two teams in a match (they find that draws are overly prevalent in the results – perhaps suggesting that teams settle for a draw after a certain amount of time in a match). The issue of dependence between goals of the two teams is revisited again and again in the academic literature: for example, Karlis and Ntzoufras¹⁹ propose an inflated bivariate Poisson model, whilst McHale and Scarf²⁰ investigate the merits of using a copula to incorporate any dependence. The assumption of the Poisson distribution itself has been challenged with some authors using alternative discrete probability distributions to model the number of goals by a team in a football match. For example, McHale and Scarf use a negative binomial distribution. Despite all having slightly different specifications, on the whole there is very little evidence to suggest a noticeable difference in performance between the large number of models based on the Poisson framework that have been proposed in the academic literature. We were supplied with a general account of the model used in FDS and it is based on this well-established Poisson framework.

Perhaps the most challenging aspect of estimating scoring rates of teams is to incorporate the time-varying nature of scoring rates in models. For example, teams have good and bad runs of form, players are injured, bought and sold, and good or bad luck can all affect the rate at which a team is expected to score in its next game. As for the underlying Poisson-type model, there are several options for including the dynamic nature of scoring rates in a model for goals. Dixon and Coles weight the results of previous matches so that more recent results have a greater influence on the estimated rate parameter for the next game than results further in the past. More recently Koopman and Lit²¹ present a state-space model for estimating dynamic scoring rate parameters.

The FDS pre-match model uses Elo updating of rate parameters. This is a viable and reasonable modelling option to adopt. Arpad Elo, a Hungarian-born American physics professor and keen chess player, developed the Elo system to produce ratings for chess²² and his work has since been adapted to the contexts of several different sports such as tennis and indeed football. The idea is simple: the rate at which a team is expected to score is updated after each match. The

¹⁷ M.J. Maher, 'Modelling association football scores', *Statistica Neerlandica*, 1982, pp. 109-118.

¹⁸ M.J. Dixon and S.G. Coles, 'Modelling association football scores and inefficiencies in the football betting market', *Applied Statistics*, 1997, pp. 265-280.

¹⁹ D. Karlis and I. Ntzoufras, 'Analysis of sports data by using bivariate Poisson models', *The Statistician*, 2003, pp. 381-393.

²⁰ I.G. McHale and P.A. Scarf, 'Modelling the dependence of goals scored by opposing teams in international soccer matches', *Statistical Modelling*, 2011, pp. 219-236.

²¹ S.J. Koopman and R. Lit, 'A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League', *Journal of the Royal Statistical Society: Series A*, 2015, pp. 167-186.

²² A.E. Elo, *The Rating of Chess Players, Past and Present*, Arco Publishing, New York, 1978.

update responds to new, more recent information about the team. The amount of adjustment to a team's scoring rate is related to the expected result of the match. For example, if a team loses a match it was expected to win, its scoring rate is updated to a lower value, whilst if a team wins a match it was expected to lose then its scoring rate is updated to a higher value than its last value.

On the basis of the description of the pre-match model supplied to us by Sportradar, it can be said that it is a perfectly conventional model properly informed by a substantial body of peer-reviewed research in the professional statistical literature. One role of the model is to serve as a back-up to reliance on the primary tool for singling out 'positives' in the first stage screen, which is the identification of substantial odds movements in the betting period leading up to a match. Substantial odds movements indicate flows of betting money on one side of a proposition, reflecting the arrival of new information on the market, which may or may not be information that the match has been manipulated. However, in some cases, bookmakers may have learned of the risk of a fix even before betting opens and adjust opening odds, so that this information is incorporated already such that no sharp odds movements are observed subsequently. Employment of the pre-match model therefore potentially draws attention to possible fixed matches which would otherwise be missed, thus improving the *sensitivity* of the testing procedure.

Estimation of the pre-match model is not universal across all competitions covered by the FDS and is therefore not part of the alerts set-up in every case. However, it is currently estimated for 103 competitions in 59 countries, so the forecasting model is employed for a high proportion of matches subject to monitoring.

Where it is employed, we found its empirical performance to be satisfactory. In use, it provides outcome-probabilities for each possible outcome in each match. These should be highly correlated with corresponding probabilities derived from observed odds in the betting market, assuming that the betting market is 'efficient'.²³ It will not be perfect correlation because the model's forecasting is based only on inputs relating to teams' past performances whereas traders on the betting market have *additional* information to exploit, such as knowledge of player absences because of suspension or injury. Market probabilities will therefore deviate from model probabilities even if 'the market' is implicitly processing information as if it were applying the same model as Sportradar.

²³ Efficiency is a concept from finance. In an efficient market, prices properly reflect all known relevant information such that prices are always 'right'. In the betting market, efficiency would imply that odds duly reflected accurately the significance of all relevant information, including the past results of the teams. There is a very extensive literature testing for efficiency of wagering markets. Generally, while identifying certain biases, sports betting markets are found to be efficient.

For a sample of 16,732 matches played over a four year period in 24 countries, model odds were compared with closing odds (the odds immediately before kick-off) recorded at Betfair. Using Pearson's correlation coefficient, the degree of correlation between the model probability of a home win and the probability of a home win according to Betfair was .952. For draws and away wins, the correlation coefficients were .839 and .946 respectively.

The high correlations between model probabilities and probabilities derived from observed Betfair odds are represented pictorially in the scatter diagrams in Figure 3.1. The model performs effectively in forecasting the market odds (expressed in probability terms) such that the predicted probabilities may be judged as a viable benchmark, to be considered by analysts when assessing alerts in the pre-match market. A discrepancy between market odds and the benchmark would be expected to be present were fixers active in the market. Of course, in the large majority of cases where there is even a large discrepancy, there will be a legitimate explanation, for example player suspensions, players tired from a midweek match, or any other reasons not captured by the statistical model. We shall note in following sections that analysts have access to relevant information that enables them to identify legitimate reasons for most apparent cases of discrepancy between model and market odds.

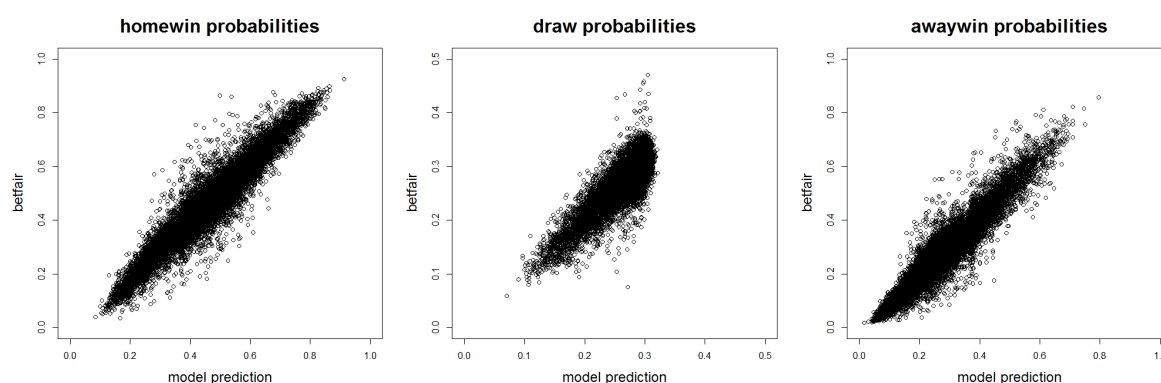


Figure 3.1. Predicted probabilities from the pre-match statistical model compared with probabilities derived from the closing odds observed on Betfair for home win (left), draw (middle) and away win (right).

3.3 The in-play model

The pre-match betting market closes once the match begins. A second type of statistical model is employed in the FDS to update outcome probabilities relating to the three principal betting markets (1X2, total goals and Asian Handicap) as events then unfold through a match. These models produce dynamic probabilities in that the estimated probabilities change as they respond to events occurring in the match in real time. For example, if a goal is scored this has an impact on the probabilities of the final results. The size of the impact will vary according to how long in the match remains.

The general description of the model employed indicated that it is similar to models developed in academic literature and by bookmakers and betting syndicates with which we are familiar. Outcome probabilities are modelled as dependent on pre-match closing odds (which will reflect the relative strengths of the two teams and the overall likelihood of goals), the current score, the time played so far, and whether one team is playing short-handed on account of one of its players having received a red card.

The impacts of red cards, goals and the passing of time on expectations of subsequent scoring rates are estimated from a large data base of historic matches. Naturally the model needs also to incorporate information regarding the scoring characteristics of the two teams playing in the particular match. Sportradar uses the prices from three markets (the 1X2, Asian handicap and total goals markets) to ‘backward engineer’ the implied odds from the in-play model (calculated at minute 0) to be equal to, or as close as possible to, the pre-match closing odds observed in the betting market.

In principle, it would be possible to account for the strengths of the two teams by using a statistical model based on their past results similar to the pre-match model outlined above. However, a statistical model based on past performances of the teams cannot allow for idiosyncratic circumstances such as absence of key players or the state of the pitch. An ‘efficient’ betting market will however factor in the influence of this extra information and it is therefore preferred to *calibrate* the in-play model using market odds (in this case pre-match closing market odds are used).

Of course, there are many platforms on which market odds are observed. For the purpose of the calibration of the in-play model, Sportradar uses the average odds from three very large Asian bookmakers: MAXbet, SBOBET and 188bet. These bookmakers operate in the most liquid betting markets where odds are most likely to be ‘efficient’. Moreover, as noted in footnote 14 above, Grant, Johnson and Oikonomidis found that odds from the Asian market were more effective predictors of match outcomes than odds from the European market.²⁴

The in-play model in practice

In the FDS, the purpose of the in-play model is to allow algorithms to create alerts where there is a significant discrepancy in the evolution of observed market odds and the evolution of outcome probabilities according to the model. This comparison is essentially between the behaviour of

²⁴ The algorithms allow for the possibility that the bookmakers specified may not offer in-play betting on a particular match: in this event, alternative platforms are specified to be used.

two sets of odds and they should be consistent with each other through the match providing that the betting market is efficient and providing that there are no relevant special circumstances in the match which are not captured by the variables in the statistical model.

We examined the evolution of model odds and market odds across many matches and found, as was to be hoped, that model odds usually tracked market odds very closely.

Figures 3.2, 3.3 and 3.4 show a typical case. They relate to a match which took place on 22nd March, 2015 between Liverpool and Manchester United.

Figure 3.2 shows the odds in the 1X2 market for an away team win, i.e. for Manchester United. The grey line (and circles) are the market odds, according to the Asian bookmaker 188bet. The yellow line (and circles) are the odds implied by the statistical model. At the start of the game the market seemed to think that the probability of a Manchester United win was slightly lower than that implied by the model (the odds for 188bet are higher than the odds for the model). This is perhaps a reflection of the volume of bets placed and the bookmaker adjusting its odds in the hope of lowering its exposure (potential for losses). After 14 minutes, Manchester United scored and both the market and the model reacted similarly and reduced the odds for a Manchester United win. In the 46th minute (the first minute of the second half), Liverpool was reduced to ten men as Steven Gerrard was shown a red card. Again, both the model and the market reacted by lowering the odds of a Manchester United win further. Liverpool scored a goal in the 69th minute and the market and model odds shifted a little in response.

There are two key points to note here. First, it is remarkable how closely the statistical model and the market, as represented by this bookmaker, quantify the probability of an away win²⁵. Second, it is equally remarkable how the market and the model agree, in terms of the magnitude and direction of the odds movements, on the effects of goals and red cards on the probability of Manchester United winning the game.

²⁵ The apparent discrepancy at the beginning of the match relates to a difference between the degree of over-round (vigourish) in the market on the particular match and the over-round artificially applied to the raw model probabilities so as to mimic the market odds for an average match with average over-round.

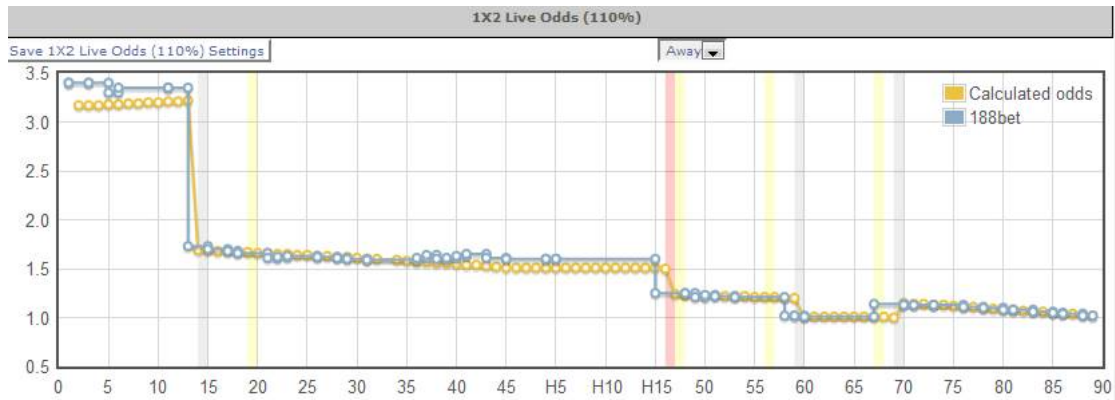


Figure 3.2: Liverpool vs Manchester United, English Premier League match, 22nd March 2015. In-play market odds and model implied odds for Manchester United win ("Away").

For the same match, Figure 3.3 shows the market odds and model-implied odds for the Asian handicap market. Again, there are some very small discrepancies but on the whole the two are closely matched. Similarly Figure 3.4 shows the evolution of market odds and model implied odds for the total goals market and as in the other two cases, there is a clear agreement.

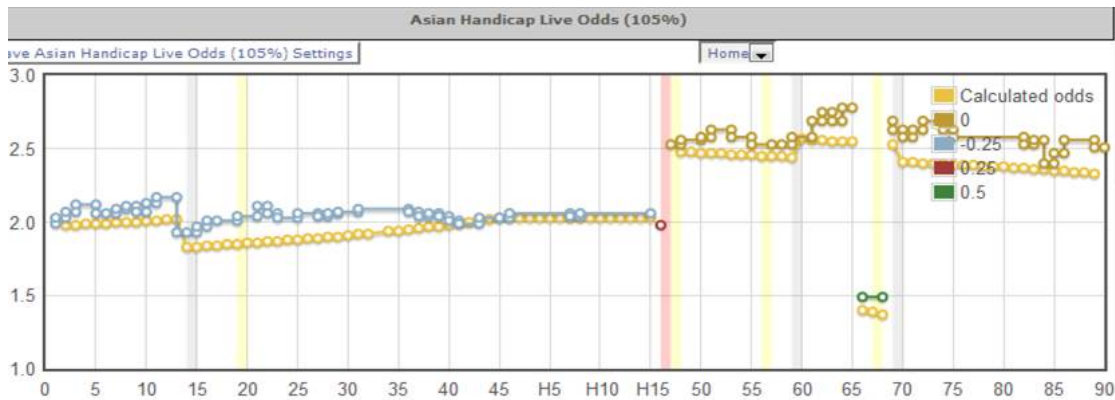


Figure 3.3: Liverpool vs Manchester United, English Premier League match, 22nd March 2015. Asian Handicap market odds and model implied odds.

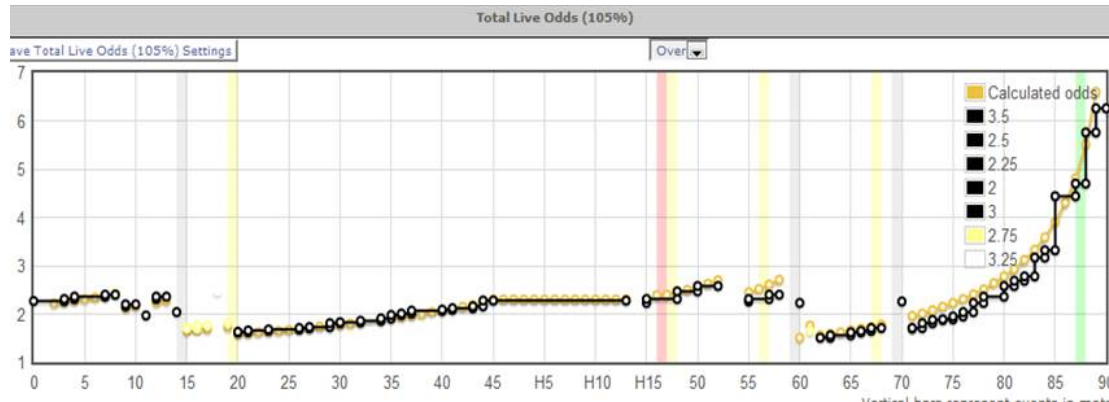


Figure 3.4: Liverpool vs Manchester United, English Premier League match, 22nd March 2015. In-play market odds and model implied odds for the total goals market.

That model implied odds and market odds track each other so closely through a high proportion of matches is consistent with the joint hypothesis that (i) the football betting market is efficient and (ii) the statistical model is working well to generate ‘expected’ odds which are quite precise benchmarks of where odds would be expected to be. The pattern is indicative that the statistical model used in-play is fully fit for purpose.

For a more formal assessment of the in-play model, we conducted an experiment which we will now describe.

Testing the efficacy of the in-play model

Testing in-play models is not straightforward. In the academic literature, the usual ‘test’ for an in-play model is to compare its estimated probabilities with those of the betting market. We have seen that, typically, the probabilities generated by the in-play model and by the market are indeed almost identical. Thus, a case might be made that formal testing of the accuracy of the model is redundant. However, whereas the first model embedded in the FDS- the pre-match model- performs only a supplementary role in identifying potentially suspicious matches (for pre-match markets, primary reliance is on identifying significant shifts in odds), the in-play model plays an absolutely pivotal role. The primary criterion for triggering further investigation of a match because of anomalous betting patterns during the game is that there is deviation between the odds implied by the statistical model and the observed odds in the market. Given this central role played by the in-play model, we therefore erred on the side of caution and subjected the in-play model to formal testing of its performance.

The in-play model predicts in real time the probability of each possible outcome of a match conditional on the odds at the start of the match, the number of goals for each team so far, the

number of red cards for each team so far, and the number of minutes remaining. The idea of the test was to look at a sample of leagues over a past period and identify all matches where a given scenario prevailed at a given time in the game. For example, one might search for all matches where the half-time score was 1-0 in favour of the home team and where the two sides had had similar odds at the start of the match and neither had received a red card. For this set of matches and the scenario specified, the probabilities generated by the model are calculated. It might be that the average probability of the final score being 2-0 given the scenario specified was $x\%$. If the model is not subject to systematic error, then the observed proportion of matches in that set which actually ended 2-0 should be close to $x\%$. If this procedure is repeated for several different scenarios, and the predicted and observed proportions are always “close”, then this would be grounds for being confident in the performance of the model. Statistical theory provides a framework for conducting a formal test and allows quantification of how confident one can be in the model.

Full details of the formal test we conducted are presented in Appendix C. This notes which leagues were used to generate the sample and sets out the ‘scenarios’ we specified. The test results presented in the Appendix indicate strongly that the in-play model performs very satisfactorily as a probabilistic forecasting model of final match outcomes. This indicates that the technical specification of the model is sound and that the information/data it uses is accurate and appropriate. It is therefore a reliable basis for identifying betting market anomalies which are potentially indicative of manipulation of the match.

4 THE ALERT PROCESS: STAGE 1 IN THE FDS

4.1 The triggering of alerts pre-match

The general approach

As noted above, the FDS is designed implicitly to follow best practice for circumstances in which both sensitivity and specificity matter. A first stage in the screening process emphasises sensitivity, i.e. great weight is placed on the importance of including all true cases in the population in the set of cases to be subject to further scrutiny. At the second stage (divided into two parts in the case of the FDS), the emphasis is on weeding out likely false positives from the first stage. This should produce high specificity at the end of the process such that there is a high probability that matches finally declared as suspicious are indeed true cases of manipulation of a match.

The first stage is conducted during the betting period running up to a match and during the match itself. This stage is entirely automated and algorithm driven.

In the pre-match betting market, the algorithms are designed such that the primary criterion for defining a positive test result (creating an *alert*) is that an unusually large odds change is observed on a single betting platform. An issue of course is how large a change has to be to be considered ‘unusually large’. Particularly in the case of Asian bookmakers, which typically adopt a book-balancing model, there will naturally be variations in odds almost continuously as new bets arrive even if there is no new information in the market. For example, Grant, Johnson and Oikonomidis²⁶ collected odds data (1X2) from the SBOBET website for 2,132 matches in the top six European leagues, played in 2012-3. Their web crawlers scraped the data for each match at eight defined points in the betting period, ranging from 24 hours to one second before kick-off. For each match, they noted how many times the odds had changed between one time point and the next. There were seven opportunities in each match for the odds to change. They found that the mean number of changes observed per match was 5.36. In other words, between any two time points, it is far more likely that odds will change than that odds will remain the same.²⁷

There is therefore likely to be a lot of ‘noise’ in the data being scrutinised. However, Grant, Johnson and Oikonomidis found that there was strong ‘signal’ amid the noise. When they added

²⁶ See footnote 14 above.

²⁷ They carried out the same exercise for the UK bookmaker, Ladbrokes, and noted the mean number of odds changes per match as only 0.795. This is consistent with their claim that European bookmakers adhere to a position-taking rather than a book-balancing model. It follows that a screen based on observing odds changes will trigger alerts more often for Asian than for European bookmakers.

odds change during the betting period to opening odds in a match outcome forecasting model, it was strongly significant. This implies that net flows of money into the market carried important information relevant to match outcome that had not been available in the market when it opened. Informed traders must therefore be active in moving the market towards efficiency. Of course, these traders will typically be processing (accurately) legitimate new information and acting on it to seek profit. But some of the traders will be acting on insider information including knowledge that an attempt will be made to manipulate the match.

This is important evidence supporting the concept underpinning the FDS: odds changes are signals of new information being acted upon and it is therefore the right thing to look among the matches where odds changes are significant if the manipulated ones are to be found. But, because the information flows are often from legitimate traders, many false positives will be liable to be declared in this first stage of the FDS process.

Fixers (and other parties who have become aware of a planned fix) are in the possession of very strong information and would therefore be expected to wager relatively large sums, making for relatively large changes in odds. It is therefore appropriate to consider further only cases where large changes in odds are observed, to contain the number of false positives. It is also practically necessary to impose a threshold to eliminate cases with only ‘small’ changes in odds. Otherwise, with so much noise in the data, a large majority of matches would be progressed to the second stage and, for most of them, no discernible reason for an odds change would be evident.

Criteria for creation of an alert

The thresholds (or cut-offs) in the FDS for defining abnormally large odds changes are set in terms of derivations of the statistic *netwin*, which is arithmetically the same as fractional odds. This appeared to us sensible. Bookmakers typically quote odds on football using the decimal odds format. For example, 1.50 means that a winning bettor on a one unit stake would be entitled to collect 1.50 money units. But this includes the return of his stake. The presence of a fixed component of one unit (stake) in the decimal odds makes comparisons across odds movements in different odds ranges problematic.

To illustrate, consider two possible odds changes.

In case 1, the quoted decimal odd changes from 1.50 to 1.20. This is a 20% change in the decimal odds.

In case 2, the quoted decimal odd changes from 5.00 to 4.00. This is also a 20% change in the decimal odds.

If the FDS defined odds changes with respect to decimal odds, the two cases would be treated as equally significant.

But, in fact, the market is making a much bigger revision of probabilities that the outcome will occur in case 1 than in case 2. In probability-odds terms, the first change is from .667 to .845 (a shift of about 18 percentage points) but the second change is only from .200 to .250 (a shift of 5 percentage points). The importance of the information driving the odds changes is therefore likely to be much more significant in case 1 than in case 2 even though the absolute change in decimal odds is greater in case 2 and the proportionate change in decimal odds is equal in the two cases. Defining odds changes by reference to decimal odds would therefore yield incorrect ranking of potential integrity risk across matches.

Use of *netwin* (fractional odds) resolves the issue. In case 1, *netwin* changes from 0.50 (1/2) to 0.20 (1/5), a variation of 60%. In case 2, *netwin* changes from 4.00 (4/1) to 3.00 (3/1), a variation of only 25%. Use of *netwin* therefore correctly signals that the market movement in case 1 is much more worthy of attention than that in case 2.

An alert is triggered whenever the difference between the highest and lowest odds observed at a single bookmaker in the betting period to date is sufficiently large in percentage terms. As noted, difference is defined essentially by reference to *netwin*; but further refinements are made in reaching the statistic finally built into the FDS for triggering alerts. This statistic is *adjusted netwin change %*. The formula incorporates an ‘exponential part’ which reduces the significance attributed to changes in longer odds ranges and an ‘additive part’ which ensures that large proportionate changes in tiny odds (for example from 1.02 to 1.01 in decimal odds terms) are not treated as seriously as an unadjusted formula would dictate.

The same principle of focusing on *adjusted netwin change %* applies in monitoring of the Asian handicap market as in monitoring of the 1X2 market. The algorithms search for movements above a specified threshold in the odds quoted for a single spread (handicap). But an additional criterion for creating an alert is also used. This focuses on odds movements during the two hours preceding a match, believed to be a period particularly favoured for fixing-related betting activity.

In the over-under market on total goals, the criterion for an alert is constructed to similar principles but is adapted to take into account that both odds and spread changes will frequently be observed when the market is reacting to heavy flows of new money. For example, in the event of a large inflow of money indicating support for the proposition that a match will yield “over” x goals, the bookmaker may change the value of x rather than just change odds within the spread. The FDS converts each odds set (spread, odds) to a single statistic, *expected number of total goals*. In this market, an alert is created whenever the change in the statistic exceeds a pre-set threshold. The statistic this time is in units of goals and changes are expressed in terms of an absolute number rather than as a percentage. The same approach is adopted when assessing variation in the market’s expectation of goal supremacy.

Additional criteria for alerts

Although pre-match monitoring is based heavily on observing variations in each bookmaker’s quoted odds, algorithms search also for other bookmaker behaviour suggestive of bookmakers being concerned over a particular match. Withdrawal of a match from the market is one such sign and an alert is created if the number of operators which have removed their odds offer is sufficiently high relative to the number offering pre-match odds on the fixture. The precise statistic used for this category of alert is given by

$$(\# \text{ of removed bookmakers}-1) / (\# \text{ of bookmakers offering odds}-1).$$

The subtraction of 1 from each of the numerator and denominator is intended to give a greater indicator of a problem where, for example, 5 of 50 operators have withdrawn compared with 1 of 5.

In the Asian handicap market, unusual skew in the odds for a given spread is also taken as evidence that bookmakers themselves have become concerned over the particular match. Usually, odds will be not far from symmetric, for example decimal odds of 1.80 and 2.00 on the two potential outcomes. If there is heavy support for one team at these odds because new information has become available, the bookmaker has the option to change the spread (handicap). If instead, the bookmaker leaves the spread unchanged but reduces odds for the now more favoured team to an extreme level, such as 1.20, this is a sign that the bookmaker does not want to accept any more liabilities against that outcome. If the bookmaker thought that the weight of money entering the market reflected legitimate news, the spread could be changed instead because there would be no commercial reason to deter new bets given that the terms of the bet would now properly reflect news in the public domain. Offering uncompetitive terms on an outcome that is suspected to be arranged already is an alternative to closing the market altogether. The FDS creates alerts when Asian Handicap odds are in uncompetitive territory defined by a threshold.

The betting exchange, Betfair, is unique in the set of betting platforms observed by the FDS in that it offers information not just on variations in odds but also on volumes of wagers transacted. While Betfair odds are monitored in the same way as those on other platforms, the FDS is also programmed to create an alert (for competitions offered by Betfair) when the volume of transactions exceeds a specified amount in absolute value *and* that value exceeds the mean turnover for a match in that competition by more than a specified proportion. Of course, if a match generated, say, eight times the normal turnover for the competition, this would not necessarily imply that anything untoward had occurred. There are many possible explanations apart from criminals engaging in a fix. For example, there may be few other football matches taking place that day or the match may be scheduled to be televised. Nevertheless, abnormal volume of betting is a characteristic of manipulated matches and the FDS aims to identify all such cases at this first stage, so that analysts may consider whether there is in fact a reasonable explanation.

Finally, for those leagues for which the pre-match statistical model is employed (see Section 2 above), an alert is created when there is more than a specified degree of discrepancy between the ‘fair (1X2) odds’ according to the statistical model and the odds observed at a bookmaker. This category of alert serves a useful function since it draws attention to cases where there is a possibility that bookmakers had learned of a possible fix (for example, through rumour) and have already factored the elevated integrity risk into quoted odds. As with other alerts, it is to be expected that there will usually be a legitimate sporting explanation; but the philosophy in the design of the FDS is to avoid missing cases of manipulation at stage 1 even at the cost of having to review a relatively large number of matches at stage 2.

Setting thresholds for alerts (‘configuration’)

We have described the principal alert categories employed in monitoring of pre-match betting markets. But these are operationalised only when thresholds are set. For example, the FDS identifies when there is a change in odds (as measured by *adjusted netwin change %*) above a certain level. But what level is to be defined as the boundary beyond which betting is to be considered suspicious?

In fact, three thresholds are set for each alert category in order to distinguish between different degrees of deviation from normal patterns of betting activity. Green, yellow and red (alternatively termed Level-0, Level-1 and Level-2) alerts refer to increasingly severe cases of abnormal activity. Green alerts are logged on to the system for information but do not require specific action by analysts at the next stage of the FDS.²⁸ Yellow and red alerts require explicit

²⁸ Green alerts may be used by an analyst when forming a judgement about hotlisting. For example, yellow alerts in-play may be scrutinised more closely if there have been green alerts pre-match (criminals make most of their profit in-play but may place modest bets in the pre-match market as well; pre-match they just dip a toe into the water

review by an analyst: it is for matches where yellow and/ or red alerts have been sounded that the analyst must decide whether or not to hotlist. The distinction between yellow and red alerts is used internally to provide analysts with additional guidance.

Different green, yellow and red thresholds are set according to the competition in which the subject match is played. This reflects that the degree of liquidity in the betting market varies hugely between leagues. In leagues which attract limited betting interest, say those of Armenia or Iceland, relatively small wagers may shift odds significantly. In a highly liquid betting market, such as that on a match in the English Premier League, the same wager may well not shift odds at all. This makes odds naturally more volatile in leagues which attract less interest and in such leagues it is judged necessary to set higher thresholds for determining whether odds changes are unusual. Otherwise the thresholds would provoke too frequent an incidence of alerts in these competitions.

When setting thresholds, the FDS distinguishes between three different tiers of competition. Level 1 competitions comprise the UEFA Champions League, the UEFA Europa League and the top divisions in England, France, Germany, Italy, Netherlands, Portugal, Scotland and Spain (and additionally the second divisions in the national structures of England, Germany and Scotland). All these competitions attract very high betting volumes and thresholds are set relatively low because even a small change in odds may reflect a large amount of additional wagering on a particular outcome, which merits examination. Examples of Level 2 competitions include UEFA international youth tournaments, the qualifying rounds for the UEFA Champions League and the UEFA Europa League, the top domestic divisions in Austria, Belgium, Czech Republic, Russia, Sweden, Switzerland and Turkey and the second division in Italy. Most other European leagues are assigned Level 3 status. Lower status of a league in the FDS signifies lower betting interest and hence more apparently random volatility in odds and a need to be more conservative when setting thresholds.

To illustrate from the 1X2 market, the *adjusted netwin change* statistic calculated by the FDS has to reach 20%/ 30%/ 50% for green/ yellow/ red alerts to be created where the match is from a Level 1 competition whereas the thresholds are set at 25%/ 55%/ 85% for Level 2 matches and at 33%/ 67%/ 100% for Level 3 matches.

because they do not wish to attract attention to a team being a good bet at current odds). Green alerts are also logged in individual player profiles. The FDS holds some 260,000 sets of player records which will include recording for any match whether any alerts had been created. Sometimes patterns may be discerned of a player having played in matches with alerts at several previous clubs.

Similar distinctions are made for other alert categories, with the value of the threshold specified varying according to level of alert (green, yellow or red) and level of competition Levels 1, 2, 3).²⁹

We inspected internal documents setting out the *alert configuration* in terms of the thresholds set at each for all categories of alert and for all levels of competition. We note that the thresholds have been defined and refined over time using the experience in betting markets of the Sportradar staff and explicitly in response to feedback from the bookmaking industry. The choice of thresholds represent a consensus and the thresholds appeared ‘reasonable’ to us also. A more formal assessment will be offered after describing the alerts process employed once the football match has begun.

4.2 The triggering of alerts in-play

Whereas monitoring of the pre-match betting markets relies mainly on selecting matches for further examination based on odds movements observed in the market, the process of detecting potentially suspicious betting activity in the in-play market is driven by a comparison between current odds and ‘calculated odds’ from the statistical model described in Section 3.3 above.

In that Section, Figures 3.2-3.4 illustrated how closely observed odds track calculated odds in each betting market as events unfold in a typical match. Screening for potentially suspicious activity at this stage involves identifying cases where the evolution of odds shows a sharp divergence from the evolution predicted from the statistical model.

Figure 4.1 relates to an example, for the Asian handicap market, of a match where odds behaved apparently perversely. This was a match highlighted by the alerts process for which

²⁹ Thresholds for the alert category focused on Betfair 1X2 turnover are the same for all levels of competition since the statistic produced by the FDS already measures turnover *relative to* the average for the particular competition or tournament. Green, yellow and red alerts are triggered when turnover exceeds 2, 5 or 10 times the league average.

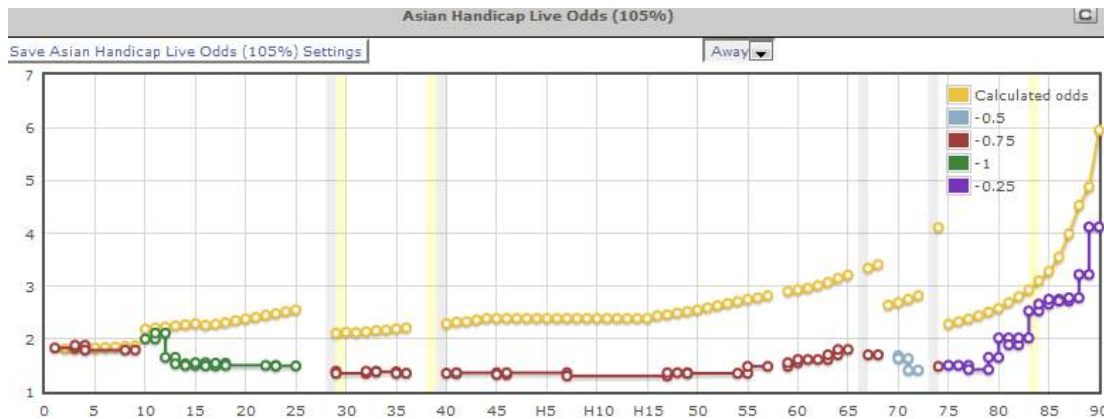


Figure 4.1: Market Asian Handicap odds and model implied odds for a match that was identified by the FDS as suspicious.

it was decided subsequently that the match should be classified as suspicious. Later, it became one of a set of matches involving the Australian club Southern Stars that were used as evidence in a criminal prosecution and for which those involved were convicted. This can therefore be considered a proven case.

The graph is for the Asian handicap market that the away team would win by more than a certain number of goals. In the early stages of the match, the model and the market odds are almost identical. This suggests that the strategy of the fixers was not to bet (at least in significant volume) before the match took place because this would alert other traders, setting up market movements which would impede the fixers' earning of maximum gains in the in-play market. However, at around minute 12, the market odds move dramatically away from the model odds. No goal was scored, nor was there any other event of significance in the game. The odds movement suggests that the weight of money in the market believed that the away team was likely to win by a large margin. The away team scores in the 28th minute and the model responds with lower odds for a large margin of victory. However, the market odds remain much lower than the model odds, implying that the away team was likely to score still more goals. This pattern continues: another goal in the 39th minute is followed by goals in the 66th and 73rd minutes making the score 0-4. Towards the end of the match the model odds and the market odds begin to converge again. This is because the fix had happened: the away team was to win by at least 4 goals. What is key here is that knowledge of the fix was being reflected in the volume bet (which is not observed) and betting volume was inducing variation in odds (which is observed). Where bookmakers follow a book-balancing business model, the odds movements reflect weight of money and here there was heavy trading (a burst of activity) by 'informed' traders who knew in advance what the course of the match would be because they had 'bought' the match.

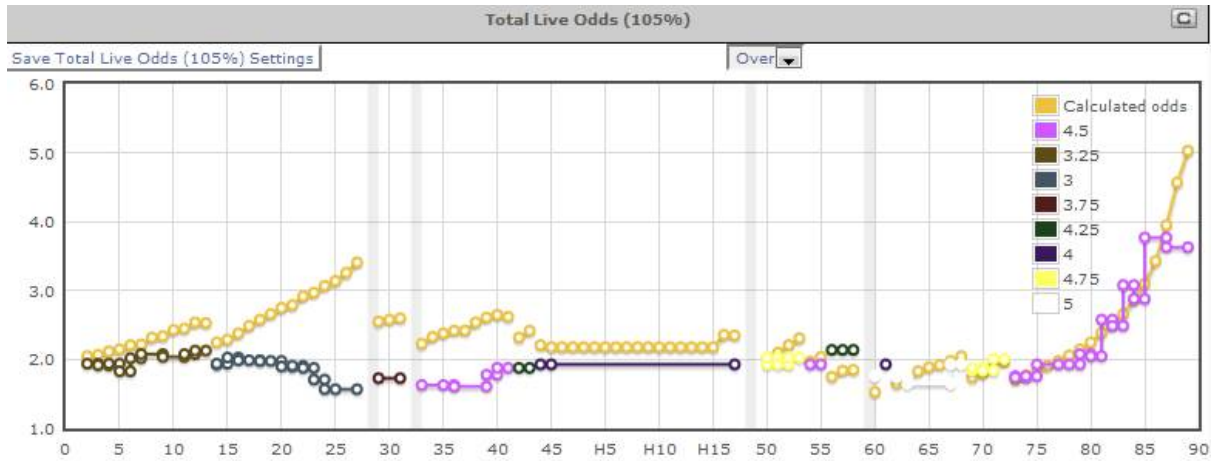


Figure 4.2: Market Total Goals odds and model implied odds for a match that was identified by the FDS as suspicious.

Figure 4.2 shows the evolution of the market on total goals for another Southern Stars match. By design the market odds and model implied odds are almost identical at the start of the match. Again they follow a similar trajectory for the opening minutes of the game. At around minute 13 the bookmaker changes the spread from over-under 3.25 goals to over-under 3 goals. Both bookmaker and model odds adjust accordingly. However, from around minute 15 the market odds begin to behave perversely. Despite there being no goals scored, and there being less remaining time for goals to be scored as the match clock ticks down, the market implied probability for more than 3 goals increases (market odds decrease). Meanwhile, the model implied probability behaves as one would expect, i.e. the implied probability decreases because there has been more passage of time without a goal. It is for such anomalous divergence between the evolution of the odds as they are observed and the evolution of the odds expected according to the model that FDS searches during in-play monitoring.

Alerts are triggered whenever there is sufficient divergence between market odds and ‘calculated odds’. In these examples, the matches were fixed but there will be other cases where there are legitimate reasons for any discrepancies observed. The statistical model accounts for major events, namely the occurrence of goals and red cards, but does not have as inputs other potentially important developments in the match. For example, suppose a key player leaves the field for medical treatment and is not substituted because the coach hopes he might return to the field later. Now that team is playing short-handed and there is a chance that it will have lost its influential player for the rest of the game. The market will process this information and odds will shift, moving them away from model odds because nothing included in the model has changed. The observed disparity this time would have an objectively valid explanation. Stage 1 screening therefore generates false positives which will have to be filtered out later in the FDS process.

As with pre-match monitoring, thresholds have to be set. The alert configuration defines cases where the discrepancy between observed and model odds is sufficiently great for the match to be selected for second-stage screening.³⁰ A ‘higher’ threshold would generate fewer stage 1 false positives but at the cost of raising the risk of missing cases where the match has been fixed.

4.3 Are the thresholds set appropriately?

Clearly it could be argued that the specification of thresholds or cut-offs must always be arbitrary to some extent. Nevertheless, we were able to satisfy ourselves that there was systematic, objective evidence to support our initial impression, that the chosen levels for thresholds seemed ‘reasonable’. Indeed, if anything, the thresholds could be said to have been set conservatively in that a high proportion of matches create yellow or red alerts and are therefore scrutinised further by analysts at Stage 2 of the FDS.

We obtained data concerning frequency of alerts across all matches monitored by the FDS during the last full football year (the twelve months to July, 2014).

Table 4.1. Matches monitored by the FDS, August 1, 2013 to July 31, 2014

number of matches monitored		45569
number of matches with yellow/red alert(s)		15129
<i>of which yellow alert(s) only</i>	<i>10601</i>	
<i>at least 1 red alert</i>	<i>4528</i>	
number of matches subsequently hotlisted		1203
number of matches subsequently escalated		291
<i>of which yellow alert(s) only</i>	<i>4</i>	
<i>at least 1 red alert</i>	<i>287</i>	

The data (Table 4.1) indicate that yellow or red alerts were created in close to one-third (33.2%) of all matches. In 23.3% of matches, the alert(s) were only at ‘yellow’ level but in 9.9% of all matches alerts included at least one at the level ‘red’. That 33.2% of matches are classified as ‘positive’ cases by the automated stage 1 screen appears, intuitively, to confirm that the net is being cast very widely at this stage. This is consistent with adherence to the academic

³⁰ Again, as with pre-match monitoring, there are additional alert categories, such as withdrawal of odds by more than a specified proportion of bookmakers which had been offering odds.

recommendation that the design of the first-stage of the screening process should emphasise sensitivity (priority to be given to include all true cases in the set of cases progressed to Stage 2).

Given the emphasis on sensitivity at this stage, should the thresholds be set lower to draw in yet more cases? Table 4.1 shows how many matches of those deemed ‘suspicious’ at Stage 1 were subsequently hotlisted (i.e. still considered suspicious following a review by an analyst and sent for further scrutiny after more information had been obtained). It also shows how many of the hotlisted matches were later ‘escalated’ (i.e. finally declared suspicious by the FDS, to be reported to the client organisation). It will be noted that a low proportion (in fact, just 1.9%) of matches identified by the algorithms as potentially suspicious are eventually classified as suspicious at the end of the whole FDS process. This is consistent with an emphasis on specificity- i.e. analysts, when making decisions about the classification at the end of the FDS process, are cautious about labelling a match as suspicious.

The large majority of matches (all except four) which were eventually classified as suspicious had attracted at least one red alert at stage 1 of the FDS. Calculations from the data in Table 4.1 show that 6.3% of matches with red alerts were eventually reported to the client as suspicious. However, of matches which had initially generated only yellow alert(s), just 0.04% (four cases) were eventually escalated to the final classification of ‘suspicious match’.

Therefore, in very few cases indeed did analysts conclude that there was sufficient evidence to support reporting a match as suspicious if it had had just yellow alerts. We established also that this was true in earlier years before 2013-14. When we obtained a list of all matches escalated over 2010-13, there were only eleven games in three years which had been classified in the end as suspicious without at least one red alert in the system.

Our conclusion is that there would be no realistic prospect of increasing the sensitivity of the overall screening system by adjusting the settings for thresholds or the protocols for using them. At present, the effective threshold for determining that there should be a further review of a match is defined by the yellow threshold. But in extremely few yellow cases is there enough evidence ultimately to label the match as suspicious. The prospect that analysing further cases with still lower levels of abnormality in the betting market (such as might, for example, trigger a green alert) would lead to an increase in the number of matches robustly identified as suspicious is therefore unlikely.³¹ We are therefore minded not to recommend any changes in the levels of the thresholds built into the alert configuration.

³¹ Referring more matches for further analysis would also have a cost in that it would inevitably lead to analysts having less time to review each match passed on from Stage 1.

5 STAGE 2 IN THE FDS: HOTLISTING AND ESCALATION

5.1 Introduction

In the academic literature, a screening procedure applied in a context where importance is attached to both sensitivity and specificity is recommended to be designed to comprise two stages. In stage 1, low thresholds (cut-offs) should be applied such that the net is cast widely in order to ensure that few true cases are missed. At stage 2, cases progressed from stage 1 are investigated further to filter out likely false positives from stage 1. The emphasis at the second stage is on assuring high specificity in the final classification of cases which have been included in screening.

In the case of the FDS, stage 2 is broken down into two parts. Analysts receive automated alerts from Stage 1 during the pre-match and in-play betting periods and, using information available at the time, determine whether there is an adequate explanation for the apparent anomalies captured by the algorithms embedded in the stage 1 process. If the analyst is not satisfied that the case can at this time be classified as a negative, and a supervisor agrees, he keeps it active in the system for further consideration once all information has been checked and additional information has been assembled (post-match).

A decision to refer the case to the second part of stage 2 is termed *hotlisting*. The second part of stage 2, where these hotlisted matches are subject to further review, is termed the *escalation* process. During this process, a decision is taken over whether the case is finally to be classified as a positive or a negative by the FDS.³² Positives are reported to clients as suspicious matches.

Sportradar provided us with a flow chart describing procedures (which are ISO-certified) during stage 2 of the FDS and this is presented as Appendix D. In this chart, the caption ‘tier 1’ refers to the hotlisting part of the process and the caption ‘tier 2’ to the escalation.

This section describes and evaluates procedures followed during hotlisting (Section 5.2) and escalation (Section 5.3) in order to inform our judgement on the efficiency and efficacy of the FDS. Our evaluation is based not only on review of internal documentation and discussions with personnel but also on live observation of both processes at the London office of Sportradar.

³² Negatives are subject to a final check at a weekly *escalation review*, which considers written reports recording reasons for all decisions not to report a hotlisted match as suspicious. This weekly meeting may decide to report a match after all.

5.2 Hotlisting

Description of the process

Stage 1 of the FDS is automated and algorithms create alerts whenever irregular betting patterns above specified thresholds are observed. Matches subject to an alert in stage 1 are the ones sent on to stage 2 for further consideration. Matches are ‘sent on’ by means of an e-mail to an analyst (and are also visible in the FDS interface). This notes that an alert has been created.³³ The analyst is then required to review the alert and all the data relating to that match in the FDS. He may also check other information sources.

The analyst reviewing a case at this stage must decide whether to label the match as non-suspicious (in which case it will become a ‘negative’ result according to FDS screening) or, alternatively, to determine that it should be hotlisted and subject to further investigation later.

Most matches at this stage are in fact labelled as non-suspicious. From data we requested, the FDS was used to monitor 45569 football matches in the year to July 31, 2014. Of these, 15129 triggered yellow or red alerts, requiring matches to be reviewed by an analyst. Then, of those matches reviewed by an analyst, only 1203 (7.95%) were hotlisted. The remaining 14926 (92.05%) were deemed non-suspicious.

In each case where a match is determined to be non-suspicious, the analyst is required to log a (brief) justification of his decision. This may be based on judgement about special factors in the betting market or in or surrounding the sporting event which make it possible to interpret apparently anomalous trends detected by the FDS as quite normal after all. For example, on the betting side, market liquidity for a particular league may vary considerably according to how much other product is available that day (a league whose season overlaps with major leagues will attract more interest when the other leagues move into their off-season). Consequently a decision over whether a particular size of shift in odds is unusual must be interpreted in the light of understanding how much liquidity there is likely to be in the market for that particular match.

From our observation of analysts at work, levels of detail logged into the FDS to justify ‘negatives’ vary. Most commonly, they attribute ‘sporting reasons’ to explain away the significance of an alert. For example, there may be a discrepancy between observed opening odds and pre-match model ‘fair’ odds because the forecasting model is not informed by

³³ Multiple alerts may be generated for a single match. Indeed this is quite common because hedging and arbitrage activity following an odds movement in one market and on one platform may then induce odds movements on other markets and platforms. In the special case of fraudulent betting, markets are likely to be independently affected by inflows of money as criminals spread their bets around to maximise profit.

information such as the suspension of a key player; sharp pre-match movements in odds may occur when news emerges that a player will be absent through injury; and deviation in-play between odds-implied probabilities and ‘calculated probabilities’ may occur after a red card because the market is able to take into account, which the model probabilities cannot, the quality of the player who was sent off. These are just examples. More generally, odds movements, particularly in-play, will reflect that the market always has more information than the models on which the algorithms are based. Market expectations of further goals (conditional on the current score and red card information) may be very different if one team on that particular day is dominating the match.

The core activity of the *hotlisting* step of the FDS is for the analyst to identify any such legitimate explanations that may account for an alert or multiple alerts having been triggered for a particular match. To perform their task effectively and accurately, the analysts need to have all possible information at their disposal, and have the knowledge and experience required to interpret the information appropriately and to estimate the magnitude and direction of the effect this information would likely have on betting odds. We discuss below the extent to which the set-up at Sportradar meets each of these two criteria.

Once all available information has been collected and its potential impact on betting odds inferred, a final, judgemental decision is made by the analyst as to whether there are sufficient ‘sporting’ reasons to explain the alerts, or that the discrepancies witnessed between market and model odds are suspicious and the match should be *hotlisted* for further investigation during the escalation stage. If the match is hotlisted, the analyst records his findings and comments for use in further discussion during the escalation process described in section 5.3.

Are the analysts adequately provided with information?

We have described the sort of additional sporting information (over and above the factual data used by the FDS algorithms, on the minute of play of each goal so far, red cards, etc) which is likely to be relevant to the decision whether or not to hotlist a match. Each analyst operates from a work station where he has access to all information held in the FDS and, for independent research around a match, online access to a plethora of football, media and other websites, and the data base of Betradar, another Sportradar product.

When an analyst responds to an alert, the FDS is likely already to contain background information on the match under consideration. Sportradar maintains a network of (currently) 43 correspondents covering football in Europe, each responsible for reporting news for a particular country (occasionally two countries). They are known as ‘freelancers’. Their duties include filing news stories relevant to the competitions for which they are responsible and writing a preview of

each upcoming match.³⁴ If freelancers' work has been carried out with efficiency and thoroughness, consulting these reports will sometimes offer immediate resolution of an anomaly. For example, unexpectedly long odds against a team may reflect news such as a recent, sudden change in coaching personnel or discontent among named players. It is plausible that these sorts of developments are likely to affect market sentiment, resulting in divergence between pre-match odds and expected odds according to the pre-match statistical model.

Typically, freelancers will also file news on player injuries and suspensions. Their information is potentially valuable here because sharp pre-match movements in odds are often interpreted as a response to changes in team line-ups. But timing of market response is crucial here if an informed judgement is to be taken as to whether the changes in the team are adequate reason to dismiss the notion that betting activity presents grounds for suspicion. For example, if a player is listed by the freelancer as likely to miss the match through injury and notes that it is a long-term injury, then this would be plausible as an explanation of opening odds being different from those predicted by the model. It would not be plausible as an explanation of subsequent movements in odds large enough to trigger an alert because the market had already known with certainty that the player would be missing and should therefore have factored this into the odds from the start of betting. The sort of information freelancers supply is therefore very relevant to the task of distinguishing between alerts according to whether they reflect suspicious patterns of events. The investment in infrastructure which Sportradar makes and of which spending on freelancers is a part clearly makes a significant contribution to the FDS. At the same time, we recognise that Sportradar exercises due prudence in the sense that internal documentation we examined guides analysts to check freelancer information against external sources whenever freelancer news has been used in deciding whether or not to hotlist.

Many alerts pre-match are created when team line-ups (who is to start and who is to be a substitute) are formally announced one hour before kick-off. Team line-ups in many cases are available reasonably quickly through a variety of sources. But in considering the significance of line-ups for betting activity, the analyst must weigh how important any changes from previous team composition are. In many cases, teams fielded in previous matches are available in the system and we noted that Sportradar provides analysts with a useful tool, Compare Line-Ups (CPU), which allows them to assess whether a player omitted from the team represents a significant omission. For example, CPU may reveal that a player dropped from the team had not in fact featured in many previous matches and the implication is that he is a fringe player whose omission should not make much impact on the betting market. This aids the analyst in weighing whether any response in the betting market is proportionate or whether it is disproportionate such that one may suspect that flows of nefarious money are driving the odds.

³⁴ Evaluation of procedures at Sportradar to assure the quality of work of freelancers is included in Section 5.3 below, when we consider the escalation process. This is the point in the FDS where freelancers' work is most likely to be decisive.

Another factor to be considered is the possibility that the changes in team line-up are themselves the means of delivering a fix. If a team fielded is very different from that expected, there could of course be legitimate reasons. However, there is also the possibility that a weakened team is being fielded to help bring about a particular outcome and indeed this means may be favoured by fixers where there is an ‘excuse’ readily available (such as resting players before an important game). On the margin, the difficulty of distinguishing between legitimate and fraudulent reasons for team changes may allow some cases to get through the net (which would marginally affect the sensitivity of the FDS).³⁵

This part of the FDS requires judgemental decision-taking by analysts and it is obvious that correct judgements depend on relevant and accurate data. On the basis of our review of hotlisting procedures, we were able to conclude that analysts were provided with an impressive array of useful information and tools, and had adequate means to cross-check the veracity of that information.

Are the analysts appropriately qualified to make the judgement?

For the analysts to make sound judgements regarding whether or not to hotlist a match, we would expect them to have an appropriate level of experience and an in-depth knowledge of the whole range of betting platforms that might throw up suspicious circumstances.

To allow us to assess the experience and knowledge of the analysts, we requested and were provided with the curricula vitae of analysts from the London (UK), Hong Kong and Sydney (Australia) offices. It is our opinion that the teams have an excellent pedigree and are appropriately experienced for the task that is asked of them. The analysts have experience of financial markets and many of them have served as traders either for betting houses or on their own account as professional bettors. Collectively their experience of financial markets covers the principal sub-sectors of the betting industry: European bookmakers, Asian bookmakers, and betting exchanges (Betfair).

Regarding qualifications, the analysts have come from a remarkably broad range of academic backgrounds. Degrees had been obtained from across the globe, including leading World-class institutions such as the University of Cambridge (BSc in Classics), the University of Durham

³⁵ A further complication to situations such as described is that team changes could be made for purely sporting reasons but that, nevertheless, there is abnormal betting activity *before* clubs hand over team sheets. This is a potential pointer to the use of ‘legitimate’ insider information for betting gain. In such a case, it is the betting market rather than the match which is manipulated. In many sports, use of insider information in this way itself violates the rules.

(BSc in Geosciences), the University of New South Wales (BSc in Computer Science), the University of Warwick (BSc in Management) and the University of Hong Kong (MSc in Statistics). Other disciplines represented in the teams included Politics, History, Finance and Sports Science.

Probably more relevant when judging whether analysts are likely to be up to the tasks asked of them whilst working as part of the FDS is the issue of experience. Almost all of the analysts had gained experience in the betting industry before starting working in the FDS and, collectively, the team of analysts had more than 85 years of experience in the gambling and betting sectors.³⁶ This included experience of many types of market including exchanges (Betfair) and bookmaking (Ladbrokes, Bet365, William Hill, BlueSquare, Betway, Hong Kong Jockey Club, Singapore Pools, SamVo Hong Kong Limited, Stan James). The majority of the analysts had served as traders in their previous roles though some have experience of other responsibilities, including croupier and fraud analyst for online poker. Several of the analysts have tried their hand at being professional bettors. Outside betting, some analysts had also worked in the broader financial sector.

The analysts, then, have substantial experience of betting markets and many of them have served as traders and/or professional bettors. Collectively their experience covers the global markets of European bookmakers, Asian bookmakers, and betting exchanges (Betfair). It is our opinion then that the teams in all three locations have excellent pedigrees and are appropriately experienced for the tasks that are asked of them. Our observation of analysts at work in the London office confirmed our view that they were well-placed to interpret events in the betting market and it was clear that their knowledge of football itself tended towards the encyclopaedic.

5.3 The escalation process

Hotlisted matches are reviewed independently by analysts prior to a group discussion of the case. This discussion takes place after information on the match has been validated and requests for additional information have been answered, which is normally within 24 hours. Each discussion involves at least one supervisor and all analysts on duty in the office. Though more will usually take part in the debate, a quorum rule requires agreement from at least three analysts for a decision to be taken to classify a match as ‘suspicious’, to be reported to the client.³⁷

³⁶ Employees with other roles than analyst also have relevant experience in many cases but this is not counted in the figure, which is only for analysts.

³⁷ Most reviews take place on Monday mornings following the weekend rounds of matches. Quite often there are then 10-15 people in the office and involved in the decision (there may also be participation from other offices). It is clear that final classifications of FDS matches are very much team decisions.

Any decision to report a match can be based only on the severity of the anomalies detected in the betting market (and whether they correspond with characteristic patterns associated with match fixing) *and* on a failure to discern any innocent reason for those (apparent) anomalies. To avoid false positives, it is imperative at this stage that the information on which the identification of an anomaly has depended is subject to checks. It is equally imperative that gathering of further relevant information, which will not have been available during the match, should be undertaken in a systematic way: further information might explain away the apparent anomaly. Only with these requirements fulfilled could one be confident in the soundness of any decision taken by the group of analysts finally to classify the case as positive according to the FDS.

Analysts are required to undertake further research on their own account during and after a match subjected to hotlisting even before information has been checked or obtained by a dedicated team within Sportradar. For example, team line-ups are swiftly available for a wide range of countries and competitions on the soccerway.com website.

All the sports data from a match which has been hotlisted is checked before the escalation process begins.

We were provided with a document which listed information sources used in these checks. This information is accessed and processed by a specialist department of Sportradar located in Germany.³⁸ The document records for each competition whether match data are ‘validated’ or ‘Sportradar checked’. ‘Validated data’ are those which are checked against official sources such as league websites.³⁹ Those ‘Sportradar checked’ are compared with information from other sources listed in the document. These other sources were very heterogeneous across countries and included websites of reputable newspapers such as L’Équipe, and credible specialist websites such as espn.com There was a checklist to show which classes of information were available from these sources, for example referee name, red cards, times of substitutions, etc. For some countries, multiple sources have to be used to obtain information across several different headings.

When a match is hotlisted, it becomes a ‘requested’ match for the data team, which uses these sources to check and, if necessary, correct details of the game previously recorded on the FDS.

³⁸ The quality management system in place in the statistics operation in Germany has been validated by ISO-certification, registration number TIC 15 100 149096, effective from December 15, 2014.

³⁹ A small number of leagues (France, Germany, Lithuania, San Marino, Slovenia, Wales) provide match data directly to the FDS. For other leagues where data are classified as validated, Sportradar collects information from official League websites as listed in the internal document with which we were provided. We verified for ourselves that the websites cited did indeed include the required statistics.

It is clear that Sportradar adopts a fully comprehensive and systematic approach when seeking out information sources with which to complete and verify the data from each match.⁴⁰ In effect, before matches are considered for possible reporting, all relevant data are in place and have been verified post-match.⁴¹

In addition to checking the information which generated the initial alerts and which was considered by the analyst when deciding to hotlist, extra information on each hotlisted match is also sought, i.e., procedures systematically require the consideration of a wider range of evidence before a match can finally be classified as suspicious. The gathering of different categories of information, beyond those used at earlier steps within the FDS, may be regarded as indicative of great care to avoid false positives being declared at the end of the process. This could of course be at the cost of decreased sensitivity. Prioritising specificity appears to us entirely appropriate given the high costs to reputation which may be borne by clubs involved in fixing cases which had not in fact been true.

The input from freelancers

A specific source of additional information is Sportradar's network of freelancers. As noted above, freelancers are correspondents, who are each responsible for the FDS-monitored competitions in a particular federation (occasionally two, typically where either or both are small jurisdictions). Their routine previews of matches, presenting news which will potentially influence outcome probabilities and odds, inform the judgement of analysts when they are deciding whether or not to hotlist a match.

Now, as part of the escalation process, freelancers' input becomes potentially yet more important. When a match is hotlisted, a request is sent to the freelancer covering the relevant competition. The request is in the form of a set of specific questions requiring a response within 24 hours. The answers are potentially important in providing evidence to justify classifying a match as suspicious. It is therefore essential that the questions asked are appropriate and that there can be a high degree of confidence in the authenticity, quality and reliability of the answers. It is also important of course that responses are actually received.

⁴⁰ For a small number of competitions, routine validation/ checking is not attempted across all matches because there are no reliable sources. All these competitions were in small jurisdictions such as Gibraltar and Liechtenstein (by contrast, some small football nations such as Faroe Islands generated very adequate data). However, 'requested' matches in these competitions *will* be subject to checking of the individual cases.

⁴¹ In fact, data on all matches to which the FDS is applied are subject to this checking process, not just data from those which are hotlisted. This is because Sportradar is the data provider for a range of sports and media partners. However, matches which are to be given further consideration as potentially manipulated are 'requested' and this gives them priority such that the data will typically have been validated/ checked by the following day.

The questions asked may be varied according to the circumstances which have prompted hotlisting of the match in the first place but generally follow a set format, with about 15 questions. These relate to issues such as whether there had been any rumours about the match, whether and why regular players were missing from the match, what the degree of motivation had been for each team to win the match, what expectations had been concerning the outcome of the match and the number of goals, whether the score in each half reflected which team had played better, whether the referee had performed well and whether the teams had appeared to give maximum effort. The freelancer may also be asked to assess each team's defensive performance, to describe key periods of play (for example, in some matches the key period of interest may be the final twenty minutes of play), to relate any noteworthy incidents within the match, and (for matches with own goals) to describe them.

The freelancer is expected to respond to these questions using personal sources and print and broadcasting media, viewing film of the match where possible. Where videos of goals are available, they are to be attached to his report.

The set of questions we reviewed seemed to focus appropriately on aspects of a match which might correlate with attempts to manipulate it. For example, defensive 'errors' are known from evidence in criminal trials to be a very common means by which goals are engineered, so it is sensible specifically to ask for an assessment of defenders' performances; and manipulation often occurs late in the match to allow the passage of time to shift odds to the advantage of fixers who know what the final score is very likely to be, so it is sensible specifically to ask for a description of this phase of the game. If suspicion has been raised at another point in the game, the question can be varied accordingly. The request for film of each goal, where it can be supplied, allows analysts to review directly the authenticity of incidents which coincide with what an unusual evolution in odds might have 'predicted'.

It appeared to us that the 'right' questions were asked to allow for a richer and better-informed discussion when analysts considered a hotlisted match during the subsequent escalation process. In so far as was possible, the questions appeared to avoid the danger of being phrased in a leading way such that freelancers might be tempted to give answers which they thought Sportradar wanted to hear: none directly asked about match fixing. Further, we were advised that freelancers also receive 'dummy' questionnaires relating to unsuspected matches, in order that a pattern is not established where the freelancer assumes he is being asked to provide details of a manipulated game. The questions were also couched in straightforward English: this is highly desirable to avoid misunderstanding given that the majority of freelancers are non-native speakers.

The reports of freelancers may prove decisive at this stage. And their role is important also in the analysts' earlier decisions on when to hotlist a match. For example, the failure of a freelancer to report a relevant news item, such as the sacking of a manager, might leave the analyst unable to account for pre-match odds not reflecting what would be 'fair' odds according to the statistical model whereas the sacking may provide a plausible explanation.

Given the importance of their role, we investigated whether adequate quality assurance procedures were in place to ensure a sufficient flow of reliable and relevant information. We confirmed that there is systematic monitoring of freelancers' performances. A quarterly report is produced to evaluate each freelancer's performance, with shortcomings identified and follow-up warnings noted.

We reviewed in detail the evaluation report for the final quarter of 2014. It assessed 43 individuals who covered competitions in 48 footballing countries. Using a five-point rating for overall performance, it graded 14 as "excellent and 1 as "poor".

Evaluation of the regular work of freelancers in entering news on to the FDS was based primarily on objective statistics of the volume of information supplied but also on judgemental assessment of the relevance of the information. Commentary drew attention to weaknesses in the flow of information from certain countries and different levels of follow-up action were specified, taking into account the individual's performance in previous quarters.

Another section of the evaluation report lists the proportion of matches where the freelancer had posted a match preview. These previews are invaluable to analysts in receipt of an alert from Stage 1 of the FDS. Two individuals had unsatisfactory records, in one case sufficiently poor that a final warning was to be issued. On the other hand, it was encouraging that there were 13 correspondents who had not missed a single match. Indeed, one had not missed a match in seven consecutive quarters.

Regarding the response of freelancers to requests for post-match reports as described above, evaluation is based on questionnaires sent to FDS staff (i.e., the users) each quarter where they numerically rate a sample of freelancer responses on five separate criteria. Here, where quality of freelancer work may be critical, there was very little cause for concern. It is noted that some

freelancers had had weak performance in terms of meeting the 24 hours deadline but the users rated the contributions of freelancers as a whole very highly in terms of accuracy and quality.⁴²

Our broad conclusion was that the work of freelancers was adequately monitored and that the framework within which information was collected from them was sufficiently well-specified so as to avoid any serious risk of inconsistency in treatment across matches.

For some matches, information is also sought from scouts. Close to 40% of fixtures are observed by scouts at the stadium. Their main role is to relay objective data on timing of the start of each half and of important events (goals, etc.). However, exceptionally, where there is a gap in information, they may be asked to complete a “feedback” questionnaire post-match which will be available for the review of the match in the escalation process. This consists of more than 30 questions which are answered as either yes/ no or on a 1-5 scale. For example, questions ask for assessment of performances by management personnel (substitutions, etc.) and by each team’s goalkeeper/ defence/ midfield/ forwards, for shirt numbers of players with outstanding or very poor performances and of players injured during the match, and whether any goals/ sendings off were controversial. It could be that there would be benefits from gathering data from scouts more systematically within FDS procedures. On the other hand, it is naturally much more difficult to monitor the performances of scouts than of freelancers, simply because they are so numerous.

The decision meeting

Once all relevant information has been assembled, the match referred for further examination is considered by all available analysts/ supervisors.⁴³ We observed discussions on four suspect matches, two involving games in a domestic league and two games played as part of an international youth tournament. Of the four matches, two were finally declared positive test results. In these cases, detailed reports were then to be prepared for the integrity officer in the client organisation.

All analysts on duty participated actively in the discussions of each case. All the evidence from the betting market and sports sphere was given full and due consideration. Video of relevant incidents was considered. Participants’ comments were clearly well-informed as would be hoped given their level of expertise (noted above) in sports trading. The impression given from the review of the four matches was that analysts tended strongly towards caution about finally

⁴² Necessarily the number of requests sent to each freelancer in any quarter will vary considerably according to the competition for which they are responsible. Some will seldom receive any requests whereas it is almost a routine part of the work in some leagues where suspicious betting activity is often observed.

⁴³ This step is not at a specifically set time but rather takes place whenever the necessary inputs into the decision-taking process are in place.

labelling a match as suspicious. This impression is confirmed by the relatively low proportion of hotlisted matches (24%) escalated in the year to July 31, 2014. Where they did take that course, they also discussed whether to assign the match a yellow or red warning level. In client reports, yellow cases are referred to as likely and red cases as very likely to have been manipulated.

All the matches for which we observed the review attracted a consensus on whether it should be reported and to what level. To the independent observer, the conclusions reached seemed to follow logically and coherently from the evidence and the debate.

The report on a match classified as suspicious

The next step is for the analyst to write a report on the match for consideration by the client in the case of a positive screen finding and by the weekly escalation review in the case of a negative. A very detailed handbook sets out a framework for writing reports. We reviewed these guidelines and we also inspected a sample of fourteen match reports. Some of the match reports were related to each other by having a particular club or clubs featuring in each. Others were ‘one-off’ reports where no suspicions had been raised hitherto about the integrity of either club in the match.

The guidelines in the handbook are intended to ensure that each report is written in a way which makes it very clear to the integrity officer at the receiving organisation why a particular match is being reported as likely to have been manipulated. Suspicious betting patterns and the reason for thinking of them as suspicious are to be explained clearly for a reader presumed to lack detailed knowledge of how betting markets function. They are to be linked to events on the field. Where appropriate, cautious reference may be made to individual players whose conduct and performance appears to be congruent with the irregular betting. Attention is to be drawn to cases where one or both clubs have attracted anomalous betting in the past, particularly in previous head-to-head encounters between the two teams.

In setting out in great detail how a report and the justifications for classifying a match as suspicious should be presented, the handbook provides a form of quality control over the standard of presentation. But it should also have other beneficial effects on the output from the FDS. In any organisation, the need to follow a framework for reporting activity helps ensure that all relevant steps are built into the execution of the activity itself. In the present case, the standard of evidence required to be included in each report implies and promotes a standard for evidence-based decision-taking during the escalation process. The common framework for reporting suspicious matches should also help promote consistency of treatment across matches.

We inspected a sample of fourteen match reports, each relating to a match within the countries of UEFA bloc of countries. All of them complied with the guidelines in the handbook. The report length was typically 40 or 50 pages. The format was to present, on the first page, a sharp and concise “summary” of the reasons for regarding a match as either “likely” or “very likely” to have been manipulated. Immediately following is the “conclusions” section of the report, which essentially expands on the summary to take the reader through the case in terms, for example, of which betting markets displayed anomalies, what these anomalies implied about market expectations concerning the evolution of the match, why these expectations were to be regarded as perverse (and not explained away by sporting information) and whether these market expectations actually predicted what was to happen on the pitch. In all cases, we found the argument straightforward to follow (and indeed, we note, convincing). We found that, consistent with the guidelines for report writing, great care was shown to explain betting market phenomena in laymen’s terms without recourse to the more specialised betting jargon. Asian handicap markets can, for example, be hard for Europeans to understand because of the complexity and unfamiliarity of the product; but movements in spreads and odds were always explained in words in terms of market expectations about the match outcome and these explanations we judged both accurate and straightforward to understand. A very useful glossary of betting terms was included at the end of the report for those who wished to understand betting terms more fully and precisely.

After the “conclusions”, each report sets out the sporting context (for example, recent form of the teams, changes in line-ups from recent matches, timing of each goal and red or yellow card in the match) and then most of the remaining pages are allocated to detailed analysis of both pre-match and in-play markets. Many pages of data in both tabular and graphical form are included to support the narrative. These, for example, trace movements in odds at major bookmakers. No doubt these data pages may be hard for some users to digest but they would be invaluable as evidence in any case where the governing body or a law enforcement agency took the case further.

Reports sometimes go beyond just explaining why betting patterns, usually allied with events on the pitch, represent evidence that a match was likely or very likely to have been manipulated. They also provide implicit guidance to the integrity officer as to what lines of inquiry might be followed and even which players or officials are most likely to have been involved.

For illustration, one match showed unusual odds movements in the totals market, indicating strong flows of money in support of the proposition that there would be at least three goals. As captured by shortening odds, this net flow in favour of three or more goals continued to be strong even as nearly a quarter of the match passed without a goal being scored. At the same time, no discernible discrepancy between odds and what odds should have been (according to the statistical model) was observed in markets on which team would win. In other words, “the

market” believed there would be more goals than one might have predicted based purely on sporting data but had formed no expectation as to which team would score those goals. This suggests that fixers’ bets were based only on insider information that there would be goals. In turn, this suggests that goals were to be manufactured by the referee rather than (for example) by poor defending by one particular team. Further, the report notes that three penalties were in fact awarded in the period of play following the period of unexplained, strong betting on at least three goals. Moreover, it notes also that each of the three penalty awards had been controversial. The story as told is clearly likely to be interpreted as suggesting that the particular match had been manipulated by the referee. Naturally this information is likely to be useful to the integrity officer in deciding on how to proceed in any investigation or it may open up other areas of action (such as referee appointments) to safeguard integrity.

In other reports, attention is drawn to previous cases in which particular players or clubs featured in the current suspected match had also been involved in previous suspected matches. A history of a player being involved in suspicious matches would be of obvious interest to integrity officers and law enforcement if they were investigating a case. Repeat offending is likely to be common in the area of match fixing since players who have been corrupted in the past will continue to be used by criminals because it is safer to use them than to approach others whose susceptibility and response to an approach is unknown. Again, criminals who have paid for a fix before have knowledge with which to blackmail the player if he does not agree to take part in subsequent fixes.

The FDS maintains a database profiling about 260,000 players and their past involvement (or non-involvement) in matches which have triggered alerts. Reports draw on this database to inform users (integrity officers) of possible patterns of repeat offending by players, referees or clubs. This is a very important part of the product offered by Sportradar since it allows the FDS, in addition to identifying likely manipulated matches, also to provide guidance on where guilt might lie. At the same time, of course, the decision to classify a match as likely to have been manipulated should not itself be influenced by suspected participation in fraud in the past: patterns over time would be less informative if a positive at one time point was reason in itself for a positive to be declared at a future time point; spurious patterns could be generated. We therefore paid great attention to those reports which pointed to past suspicions regarding individuals or clubs involved in the subject match. In all such cases, we were satisfied that the case for positive classification of the subject match could be made independent of the observations in the report concerning past involvement in suspect matches.

Conclusion

Our review of the part of the FDS driven by expertise and judgement rather than by algorithms confirmed that procedures are in place to ensure that decision-taking is based on thorough and

accurate information. A very large proportion (98.1% in 2013-14) of matches classified as positive by the first, automated stage of the FDS are subsequently classified as negative once analysts have considered the circumstances of the match in either or both of the hotlisting and escalation parts of the process. From our live observations, analysts indeed show caution by being ready to accept other explanations than fraud for betting anomalies. Implicitly, they appear to sacrifice sensitivity for specificity, i.e. some cases of fraud are missed by the FDS because only cases of very striking betting anomalies which are entirely without explanation are finally escalated. In our view, the way in which the FDS is operated makes it likely to produce classifications exhibiting high specificity, i.e. the probability of a false positive in final classifications is likely to be low.

6 CASE STUDIES: THE FDS IN ACTION

6.1 Introduction

As noted at the start of this Report, it is not possible to compute numerical measures of the performance of the FDS according to the key criteria of any screen, sensitivity and specificity.

Sensitivity is essentially unknowable as any successful fraud which escapes classification as a positive according to the screen is unlikely to come to light subsequently. There can therefore be no definitive count of missed cases (false negatives). However, we have been able to show that the coverage of betting markets by the FDS is wide enough and the technical specification of the system well-constructed enough for it to be implausible that any attempted manipulation of a football match involving large bets will escape detection.⁴⁴

Specificity is also impossible to evaluate numerically to the extent that most positive results from the screen cannot later be classified as either true positives or true negatives. In many cases, further action following a report from Sportradar does not happen or does not come to light. This may be because the governing body has inadequate support from the relevant law enforcement agency, or even that it may prefer to ignore cases⁴⁵ drawn to its attention because any investigation would be disruptive, costly and potentially commercially damaging (for example, if revelations deterred sponsors or lowered spectators' confidence in the authenticity of the competition).⁴⁶ Similar issues are noted in the academic literature on anti-doping where federations may be complicit with cheats simply by neglecting to pursue positive test results. With FDS, even where a case is followed up, there is typically no transparency in terms of what has been learned from inquiries and therefore no systematic record of which FDS-suspicious matches have been regarded as true positives, false positives or not proven.

While it is not possible to quantify the performance of the FDS in terms of formal sensitivity and specificity indices, it is instructive to review some case studies of matches and groups of matches where the FDS has detected evidence of suspicious betting activity and those matches have subsequently been 'proven' as fixes (at least to the extent of police prosecution, conviction of players in court or banning of players by national or international governing bodies). We consider in this section cases from football in Australia, Austria and the Baltic states. The cases

⁴⁴ Unsophisticated small-scale fraud is less likely to come to light through the FDS. The culprits tend to bet relatively small sums with immediately local operators, with, for example, little impact on odds worldwide.

⁴⁵ E.g. in 2013 the South African Football Association lost a sponsorship deal following match fixing revelations coming to light (<http://uk.reuters.com/article/2013/10/17/uk-soccer-safrica-puma-idUKBRE99G0AK20131017>).

⁴⁶ In some jurisdictions, fixing may be bound up with control by criminal gangs, with officials intimidated from investigating suspicious matches.

from Austria and Estonia in particular have features which will allow us to draw out evidence at least very suggestive of the FDS performing to high standards of both sensitivity and specificity.

6.2 Case studies

Australia

Sportradar was directly involved in the uncovering of fixing in the Victorian Premier League, which may be regarded as second-tier soccer in Australia (of similar quality as, perhaps, sixth-tier football in a country such as England). Despite the modest status of the competition, associated betting markets are believed to be highly liquid because betting on Australian football is popular in East Asia on account of the time zone being more conducive than in Europe to betting and following the game at the same time. High liquidity makes potential profit from fraud high and there is therefore *a priori* reason to suppose that Australian soccer (particularly with its generally low wages) faces serious integrity risk.

A number of matches proved to have been manipulated by employees of the Southern Stars club. Four of its players (all from the United Kingdom) and a coach were charged by Victoria Police and convicted in the courts alongside a Malaysian national who had liaised between the players and the betting syndicate which had paid for the fixes. Although not all of the fixes had been ‘successful’, the syndicate was reported to have made an estimated AUD2m. from Southern Stars matches manipulated between July 21 and September 13, 2013.⁴⁷ Some of the bets were said to be “in the hundreds of thousands of dollars”⁴⁸, testimony to the high liquidity in the betting market even on a match at such a relatively modest level of the sport.

Investigation of the case was the direct result of pro-active monitoring of associated betting markets by Sportradar. At the time, it did not have a contract to monitor matches in that particular competition. However, analysts, who naturally keep abreast of football news, noted that a number of English sixth-tier players had been transferred to the Southern Stars club and had in common that each had played in English matches which had been declared as suspicious by the FDS. The simultaneous transfer of several such players to one club alerted analysts to the possibility that match fixing would occur at their destination club. Analysts therefore monitored Southern Stars games over a period.

⁴⁷ S. Bricknell, ‘Corruption in Australian sport’, *Trends and Issues in Criminal Justice*, no. 490, February, 2015, Australian Institute of Criminology, Australian Government.

⁴⁸ *The Australian*, September 16, 2013.

This proved to be the undoing of the conspiracy to use the Southern Stars club as a means of making fraudulent gains on the betting market⁴⁹. “Sportradar detected irregular betting patterns associated with at least five Southern Stars games, which were characterised by ‘unusually poor play’ by some of the players” (S. Bricknell, see footnote 39). The positive results from the FDS screen were reported to Football Federation Australia (the FFA) by Sportradar and the FFA put Sportradar in touch with the Victoria Police the following day. The police began ‘Operation Starlings’ which ultimately lead to court hearings in which players admitted their involvement in fixes.

We had sight of a letter to Sportradar from Graham Ashton, Acting Chief Commissioner of Victoria Police, dated November 18, 2013. It acknowledged the role of Sportradar both in identifying the presence of match fixing in the League *and* in assisting the investigation with further advice and analysis. For example, Mr. Ashton wrote: “Sportradar identified the match fixing issue within the Victorian Premier League and subsequently supported and provided specific match odds advice to Victoria Police”.

The role of the FDS was also widely acknowledged in media reports on the case and there is no doubting that the positive screen results on FDS lead directly to uncovering and successful prosecution of match fixing in Australia. Subsequently this success lead to the signing of a Memorandum of Understanding, dated April 3, 2015, between the Australian Federal Police and Sportradar, with a view to cooperation which would help law enforcement objectives to be achieved. The case is therefore illustrative of the powerful role the FDS can have in attacking match fixing when sports federations and law enforcement actively pursue reports that matches are suspicious according to the FDS. No doubt the willingness of authorities to investigate is greater when a single club has been implicated in several matches adjudged suspicious rather than just one.

It is also interesting to reflect on the earlier matches in England which the FDS had labelled as ‘suspicious’ and in which players convicted in Australia had taken part. Naturally, the court in Australia could not adjudicate on these games and they are not in that sense ‘proven’ cases. On the other hand, that players who had been involved in positive screen results in England went on to agree to fixing in Australia lends great credibility to Sportradar warnings on the earlier English matches, which we are minded to treat as (almost) proven cases.⁵⁰

⁴⁹ It later became apparent that several English players had been recruited to the Southern Stars club by an agency which was a front for match fixers: criminals identified and brought together players whom they believed to be corruptible with a view to systematically manipulating results at their new club (*The Guardian*, July 17, 2014).

⁵⁰ The (English) Football Association (F.A.) appears at the time not to have made any public response to the positive FDS test results and warnings from bookmakers about these ‘Conference’ matches other than to send a general warning to all clubs in the competition that they should remind their players and officials about their responsibilities under betting and integrity rules (*The Guardian*, March 15, 2013). This drew press criticism after the Australian

Our interpretation of events is that Sportradar took the opportunity of using positive results from the FDS which appeared not to have been possible to follow up in one football jurisdiction to *predict* later manipulation of matches in another football jurisdiction. That the predictions came to pass is highly suggestive that the earlier positive screen results were likely to have been true cases and this is encouraging as a test of the specificity of the FDS screen (i.e., from later events, cases declared as positives seem very likely to have been true positives).

Austria

An investigation by Austrian police resulted in prosecution of five players or former players and five other individuals on charges related to the manipulation of eighteen matches in the top two divisions of the national league over 2004-2013. Eight individuals were found guilty. Sentences handed down by the court included a prison term of five years for Sanel Kuljic, a well-known player who had played for the national team twenty times. Associated bets on individual matches had ranged to €300,000 and were arranged by Albanian criminals on Asian markets.⁵¹

The police operation in this case was not initiated in response to Sportradar reports. We carefully studied press reports on the case⁵² and satisfied ourselves that one of those later convicted actually approached the police alleging blackmail and intimidation by one of his co-conspirators after a failed fix. This was the trigger for the police investigation and therefore the revelation of the proven-at-law cases was exogenous (independent of the FDS). This sets up a test of the efficacy of the FDS because it is possible (for some matches) to go back to the FDS and find out whether it had in fact identified these matches as suspicious.

In fact, Austrian police went further than this and initiated what was effectively a scientific experiment to evaluate not only the sensitivity but also the specificity of the FDS.

The police as part of its inquiry provided a list of matches for which it had evidence of fixing and asked Sportradar whether each match had been regarded as suspicious at the time. But it included in the list a number of other matches for which there was no reason to suspect malpractice and which were therefore likely to have been free of corruption.

case. It seems that the F.A. blamed the unwillingness of police to investigate for their apparent lack of action (*Daily Mail*, September 6, 2014). It should be noted, however, that, since then, Britain's new National Crime Agency has shown a strong commitment to pursuing match fixing.

⁵¹ information from www.lawinsport.com and www.worldsoccer.com

⁵² for example, *The Guardian*, November 19, 2013

If the FDS screen had ‘perfect’ sensitivity, it would have issued warnings at the time for all of the matches for which the police had acquired evidence of corruption. If the FDS screen had ‘perfect’ specificity, it would not have issued warnings for any of the innocent matches.

We were provided with a written testimony (original in German, read in English translation), dated March 16, 2015, from M.A. Holzer, Head of Office at the Organised Crime Bureau in Vienna. The testimony confirms that: “all matches which were graded by Sportradar as suspicious to be manipulated for betting purposes, the evidence of match-manipulation was provided during the criminal investigations. Other unsuspected matches have been identified by Sportradar as not suspicious for match-manipulation” (from the translation of the letter provided).

On the face of it, the experiment from Austrian police appears to be consistent with perfect sensitivity and specificity. However, the statement provided is vague and does not specify how many matches were tested and what were the absolute numbers of true and false positives and true and false negatives. Without these numbers, it is not possible for us to offer formal hypothesis tests regarding the value of sensitivity and specificity indices.⁵³ On the other hand, the willingness to commit to the testimony signifies that the Austrian police was impressed by the performance of Sportradar both in identifying matches as suspicious where these were independently discovered to be corrupted and in clearing matches where actually no grounds for suspicion existed. Moreover, Austrian Police subsequently signed a Memorandum of Understanding with Sportradar, dated July 2, 2014, committing to future cooperation between the two organisations. The associated press release quoted Mr. Holzer, Head of the Organised Crime Bureau for Austria, as commenting that “we have brought Sportradar on board to ensure that we keep our finger on the pulse. Their expertise and intelligence has been, and will remain, invaluable as we protect the integrity of sport in our country”. Thus, though the evidence is somewhat informal, we are willing to accept it as at least highly suggestive that the FDS works effectively when judged against the criteria of sensitivity and specificity.

Estonia and Latvia

We reached similar conclusions from our review of a lengthy investigation of match fixing in Estonia. The affair came to public light in December, 2013, when police arrested eight players (including the League’s all-time top goal scorer) and three other individuals believed to have participated in the manipulation of seventeen matches during 2011-2012. Mainly they were matches in the top division in Estonia but included three played as part of the UEFA Europa League.⁵⁴ During 2014, the players concerned were banned by the Estonian Football Association

⁵³ We were advised that Austrian police felt unable to give numbers because the investigation of match fixing in Austrian football is still ongoing (and further matches/ players may become implicated).

⁵⁴ www.bbc.co.uk

and these bans were later extended to World bans by FIFA. At the time of writing, the players appear to have escaped criminal penalties as the Tallin Circuit Court ruled that the charges against them should be dropped. However, the Court's decision was based on technical grounds related to the lack of an explicit provision against match fixing in the Estonian criminal code. They had been tried on an inappropriate offence and the decision did not imply that there had been no match fixing.⁵⁵ From the willingness of the football authorities to impose bans, the readiness of the public prosecutor to take the case to the courts, and the content of the judgement, it would seem reasonable to consider the matches in question as 'true' cases of fixing.

We were given a copy of a letter, dated May 5, 2015, from Mihkel Uiboleht, Integrity Officer at the Estonian Football Association. It notes that FDS reports played a key role in its investigation, which lead ultimately to 26 players receiving sporting sanctions in 2014. Mr. Uiboleht makes the following especially relevant observations (our italics): "The [FDS] reports and accompanying analysis were used both to launch the investigation *and also retrospectively to confirm intelligence already gathered* which were under suspicion of match-fixing. In the vast majority of cases, the criminal intelligence corresponded to highly suspicious betting patterns recorded [by the FDS] in matches already escalated independently of our investigation".

Again, without numbers of matches involved, the statement cannot permit statistical testing of specificity. But, as with the Austrian case, the description is suggestive of 'good' specificity in that, where independent intelligence revealed a likely fixed match, it also appeared so when FDS records were consulted *ex post*.

Another of the Baltic states, Latvia, also saw arrests during 2014 in connection with match fixing schemes. In October two players and two club officials from Daugava Daugavpils (as well as four other individuals alleged to be involved in organising the fixes) were detained.⁵⁶ An official statement by UEFA⁵⁷ explicitly linked the initial investigation of the case to a report by FDS concerning irregular betting activity surrounding the club's UEFA Champions League match against a team from Sweden.

Other cases where FDS reports were acted on by the football and the civil authorities were drawn to our attention. We were presented with a written statement by Urs Kluser, Integrity Officer at UEFA. It referred to arrests or police questioning of players between December, 2014 and March, 2015, in Bulgaria, Moldova and Montenegro. These cases are clearly ongoing but the actions of law enforcement suggest that evidence gathered already (if not yet tested in court) is strong. The UEFA statement confirms that, in the Latvian case and in the three countries with

⁵⁵ www.news.err.ee (Estonian public broadcasting)

⁵⁶ www.uefa.org/protecting-the-game/integrity/news/mewsid=217235.html

⁵⁷ for example www.goal.com

more recent action by prosecutors and police, the initial trigger for inquiry was the receipt of FDS reports of suspicious matches. The matches in these cases all fall into the category of positives on the screen and likely from subsequent investigation to be true positives.

All the cases considered here represent ‘success stories’ for the FDS and demonstrate the benefit a sports federation committed against match fixing can gain when betting markets around its matches are monitored. However, we have reviewed them from the perspective of seeking clues as to the likely degrees of sensitivity and specificity of the screen. We regard specificity as particularly important in that high costs may accrue to clubs falsely accused of involvement in manipulation of matches. Details of the cases, in so far as it is possible to draw conclusions from small samples, reinforce our view that the FDS exhibits ‘good’ specificity.

7 SOME REFLECTIONS

The FDS produces output in the form of classification of matches as either suspicious or not suspicious. Whenever any system in any industry produces output, the most obvious way of assessing its efficiency and efficacy is to test the quality of that output directly. However, this is not always possible. For example, an industrial process may produce output which is high value and where quality can be confirmed only by a test involving destruction. In such cases the second best approach to assessing quality is rigorously to examine the system's constituent parts. Examining whether these parts are designed properly, whether they are reliable and accurate and whether they are used appropriately, is a means of establishing whether one can be confident in the quality of the output.

Here, our assessment of the FDS could not be informed by certain knowledge of whether the classifications generated by the process were correct. Therefore, although we were able to make some inference from case studies, our assessment generally relies on a detailed examination of the constituent parts of the FDS.

Accordingly, we broke the FDS into steps. First in the FDS, data from both sport and the betting market are assembled. The data are then examined in an automated way using algorithms from two mathematical/statistical models. We evaluated the scope and reliability of the data input, the soundness of the mathematical/statistical models, and how they were put to use in initial classification of matches as either not suspicious or worthy of further examination. Second, those matches which are to be scrutinised further are then examined by analysts with access to the data already assembled and to additional information obtained from a network of correspondents covering each football country monitored by the FDS. If a match is still considered suspicious it is reviewed again at a meeting of analysts after yet more information has been gathered and all the earlier data verified. We evaluated the decision-making process, the qualifications of the analysts and the reliability of the supplementary information employed at this step.

Generally, our audit was reassuring as to the soundness of the system at all steps and therefore there is reason to be confident in the quality of the output, the final classification of cases as either positive (suspicious) or negative (not suspicious).

This is not to say, of course, that the FDS exhibits 'perfect' sensitivity and specificity. That would not be possible. Sensitivity will be less than perfect: some cases will be missed, for a range of reasons. First, small-scale fraud may not register to the extent that flows of nefarious money will be insufficiently large to shift odds by enough to be significant amid the general noise surrounding the odds data. Second, a large proportion of matches which generate alerts from the automated system are then classified as negatives by analysts because they find sporting

reasons for apparent betting market anomalies. Sometimes these ‘sporting reasons’ may hide a fix, for example one engineered through changing team composition. On other occasions, it is inherently hard to evaluate whether an idiosyncratic item of sporting information is capable of explaining an odds anomaly of a particular size. If analysts err on the side of caution in these cases, for fear of mislabelling a match as likely to have been manipulated, then this will lower the sensitivity of the FDS. Our observation of analysts at work suggested that they are indeed cautious in their approach; but this is not a criticism as compromising on sensitivity is normal and proper when designing screens where false positives would be costly to the parties concerned. What can be said, from the case studies, is that the system has been demonstrated as having successfully detected several recent and proven attempts by criminals to exploit football and its associated betting markets. We do not doubt therefore that the FDS is contributing in a very valuable way to protecting the integrity of the sport.

The power of the FDS to protect the sport depends on Sportradar having built an impressive infrastructure to deliver its product. Construction of the infrastructure has necessitated considerable investment in both physical and human capital. The physical capital comprises technology for gathering and processing data from sport and from betting platforms around the World. The human capital comprises expertise in the form of analysts in the Sportradar organisation and a network of correspondents (freelancers) supplying them with relevant, often non-quantitative information. Our review concludes that this infrastructure of physical and human capital is used effectively.

As independent consultants, we are struck not just by the success stories of the FDS but also by the disparity between the number of positive screen results generated and the number of cases with known follow-up action by the sports competition or a law enforcement agency. Reasons for inaction are varied and may include, for example, lack of resources or a legal framework within which match fixers can be pursued. But we note that, if the reason is that sports federations typically prefer not to pursue further any reports of suspicion surrounding matches, then this necessarily undermines the capacity of the FDS to contribute to safeguarding the integrity and authenticity of sport. Most fixes are executed by athletes themselves (rather than, say, referees) and their decisions on whether or not to engage in a fix must depend partially on their perception of how likely it is that their behaviour will be both detected and punished. Athletes’ decision will therefore be informed by expectations about whether the governing body will actively investigate positive screen tests. Game theory suggests that any tendency to fail to investigate thoroughly will feedback into a greater willingness of athletes to cheat.

This nature of the problem is familiar from debate on doping in sport. Governing bodies in certain sports appear to be far from rigorous in following-up positive blood test results, presumably because they have a disincentive to do so (for example, potential loss of public

support for the sport). Rational athletes therefore dope because they know that there is low probability of any detection being followed by punishment.

Addressing this problem in both the doping and fixing spheres would involve giving governing bodies greater incentive to investigate possible integrity offences. One means of doing this would be to enforce greater transparency in the behaviour of governing bodies. For example, we would recommend to confederations that they encourage member countries to agree to a convention that they had, at a minimum, to confirm that they had considered each report from the FDS and to outline what type of response they had made.

8 CONCLUSIONS

As independent consultants, we have carried out for this Report a detailed review of Sportradar's Fraud Detection System (FDS) which is based on monitoring betting markets for anomalous activity which might indicate that a sports event has been subject to fraudulent manipulation.

We noted that the efficacy of screening for any phenomenon is conventionally judged by the extent to which the classification of cases as positives (here, fixed matches) or negatives (here, non-manipulated matches) exhibits sensitivity and specificity. Sensitivity refers to whether the screen picks up a high proportion of true cases; specificity refers to how confident one can be that cases classified as positives are true positives.

We examined in detail every component of the FDS.

- In Section 2, we examined the data input into the first, automated stage of the FDS. We found that the breadth of coverage of betting markets was very wide such that betting activity related to significant fraud was very likely to be picked up. We scrutinised the betting data and the sports data to which the algorithms for detecting fraud are applied. The betting data were free of error. The sports data were gathered from a comprehensive list of sources and were subject to robust checks to ensure accuracy.
- In Section 3, we examined the mathematical and statistical models which drive the algorithms used to identify potentially suspicious matches. They conformed to best practice in their construction and performed well when we subjected them to empirical testing.
- In Section 4, we examined the selection of criteria embedded within the algorithms for evaluating whether there were anomalous patterns of activity in betting markets and the corresponding thresholds used to define which matches needed further consideration. We found that the criteria employed were conceptually sound and allowed for possible manipulation in all the principal markets offered on football matches. Thresholds were set quite low such that a significant proportion of matches were flagged as requiring assessment in the second stage of the FDS, when analysts become involved. We presented evidence that setting thresholds any lower would be unlikely to lead to more matches being classified at the end of the FDS process as likely to have been manipulated. We therefore recommended no change in thresholds.
- In Section 5, we examined Stage 2 of the FDS where analysts decide whether the matches with betting anomalies drawn to their attention by Stage 1 are truly likely to have been manipulated. First, analysts filter out cases (a large majority) where they perceive a ready, legitimate explanation for apparent anomalies. Those still then regarded as

potentially suspicious are referred on for more detailed scrutiny, which involves group decision-taking on whether to report a match as likely to have been manipulated. Prior to this final scrutiny, all sports and betting data are checked and further relevant information obtained. In reviewing processes in these parts of the FDS, we were satisfied that the procedures for reaching a decision are rigorously set out and followed. We determined that the qualifications and collective experience of the team of analysts equip them to make reliable assessments of the evidence. They were informed by data which had been subject to appropriate checks according to systematic procedures and by appropriate additional information obtained from correspondents on the ground. We noted that only a very small proportion of matches flagged as potentially suspicious by the algorithms in Stage 1 were finally classified as likely manipulated by the analysts' team in Stage 2. This we judged to reflect a cautious attitude where sensitivity was implicitly sacrificed in favour of specificity: only matches where a compelling case could be made were in the end reported as suspicious to the relevant sports organisation.

- In Section 6, we examined some case studies relating either to matches reported as suspicious by Sportradar and subsequently verified by the legal system as manipulated or to matches independently discovered to have been manipulated. In the case of the FDS it is not possible to assign a precise numerical value to the level of specificity because many reports are not investigated further to establish finally the truth of whether manipulation has been present. However, this review of some known instances of match fixing provides evidence fully consistent with high specificity.
- Our overall conclusion from the study is that matches reported as suspicious by the FDS are very likely to have indeed been manipulated.

Appendix A: List of bookmakers monitored within the FDS

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
888	Yes	No	Direct feed
10Bet	Yes	No	Direct feed
188bet	Yes	No	Direct feed
188bet LiveOdds	No	Yes	Direct feed
188BetBU	Yes	Yes	Scraped
1xbet	Yes	No	Direct feed
5Dimes	Yes	No	Direct feed
855WinBU	Yes	Yes	Scraped
ACTTAB	Yes	No	Scraped
Admiralbet	Yes	No	Direct feed
AFB88BU	Yes	Yes	Scraped
AFB88LiveOdds	No	Yes	Direct feed
Ambassador	Yes	No	Direct feed
Balkanbet	Yes	No	Direct feed
BalkanBetLiveOdds	No	Yes	Scraped
Baltbet	Yes	No	Scraped
bet-at-home	Yes	No	Direct feed
Bet-At-HomeIt	Yes	No	Direct feed
Bet18Com	Yes	No	Scraped
Bet3000	Yes	No	Direct feed
Bet3000LiveOdds	No	Yes	Direct feed
Bet365	Yes	No	Direct feed
BetAtHomeLiveOdds	No	Yes	Direct feed
BetCafeArenaRomania	Yes	No	Direct feed
BetCity	Yes	No	Scraped
BetClic	Yes	No	Direct feed
BetClick.fr	Yes	No	Direct feed
BetClickITLOHidden	No	Yes	Direct feed
Beteasy	Yes	No	Direct feed

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
Betfair Australia	Yes	No	Direct feed
BetfairSportsbook	Yes	No	Direct feed
Betflag	Yes	No	Direct feed
Betfred	Yes	No	Direct feed
Betgun	Yes	No	Direct feed
Betinternet	Yes	No	Direct feed
BETISN	Yes	No	Direct feed
BetISNLiveOdds	No	Yes	Direct feed
BetJack	Yes	No	Scraped
Betonline	Yes	No	Scraped
Betpro	Yes	No	Direct feed
BetRedKings	Yes	No	Direct feed
Betsafe	Yes	No	Direct feed
Betsson	Yes	No	Direct feed
BetsSonLiveOdds	No	Yes	Direct feed
Better	Yes	No	Direct feed
BetVictor	Yes	No	Direct feed
BetVictorLiveOdds	No	Yes	Direct feed
Betway	Yes	No	Direct feed
Bingoal	Yes	No	Direct feed
BookmakerComAu	Yes	No	Scraped
Boylesports	Yes	No	Direct feed
bwin	Yes	No	Direct feed
Bwin.it	Yes	No	Direct feed
BwinFr	Yes	No	Direct feed
bwinLiveOdds	No	Yes	Direct feed
Casapariurilor	Yes	No	Direct feed
CashPoint	Yes	No	Direct feed
CashPointLiveOdds	No	Yes	Direct feed
CBCX	Yes	No	Direct feed
Centrebet	Yes	No	Direct feed
CentrebetLiveOdds	No	Yes	Direct feed
Centurionbet	Yes	No	Direct feed

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
CMD368LiveOdds	No	Yes	Direct feed
CMDBetBU	Yes	Yes	Direct feed
ComeOn	Yes	No	Direct feed
Coral.co.uk	Yes	No	Direct feed
Danske Spil	Yes	No	Direct feed
Digibet	Yes	No	Direct feed
DigibetLiveOdds	No	Yes	Direct feed
Doxxbet	Yes	No	Direct feed
EasternDynastyBU	Yes	Yes	Scraped
EccobetAlbania	Yes	No	Scraped
Efbet	Yes	No	Direct feed
Empire Betting	Yes	No	Scraped
Eurobet.it	Yes	No	Scraped
Eurofootball	Yes	No	Direct feed
Eurolive.al	Yes	No	Scraped
Eurolloto	Yes	No	Scraped
Expekt	Yes	No	Direct feed
Favbet	Yes	No	Direct feed
FlemingtonsportsbetCom	Yes	No	Scraped
Fonbet	Yes	No	Direct feed
Fortuna-sazky	Yes	No	Direct feed
Francaise des Jeux	Yes	No	Direct feed
Gazzabet	Yes	No	Direct feed
GenybetFr	Yes	No	Scraped
GenybetFrLiveOdds	No	Yes	Scraped
GermaniaSport	Yes	No	Direct feed
Gioco Digitale	Yes	No	Direct feed
GiocoDigitaleLiveOdds	No	Yes	Scraped
GoldBet	Yes	No	Direct feed
Goldbet.al	Yes	No	Scraped
Guts	Yes	No	Direct feed
GWBet	Yes	No	Direct feed
Hattrick	Yes	No	Direct feed

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
Hong Kong JC LiveOdds	No	Yes	Scraped
IGKbetBU	Yes	Yes	Scraped
Inteltek	Yes	No	Direct feed
Intertops	Yes	No	Direct feed
Interwetten	Yes	No	Direct feed
Intralot	Yes	No	Direct feed
Izibet	Yes	No	Direct feed
Ladbrokes	Yes	No	Direct feed
Ladbrokes Italy	Yes	No	Scraped
Ladbrokes LiveOdds	No	Yes	Direct feed
Leon Bets	Yes	No	Direct feed
LigaStavok	Yes	No	Direct feed
LiveBet365	No	Yes	Direct feed
LiveBwin	No	Yes	Direct feed
LiveBwinIt	No	Yes	Direct feed
LiveInterwetten	No	Yes	Direct feed
Loterija	Yes	No	Direct feed
LottomaticaLiveOdds	No	Yes	Scraped
LuckiaLiveOdds	No	Yes	Direct feed
LuckystreamLiveOdds	No	Yes	Direct feed
Lutrija	Yes	No	Scraped
Macau Slot	Yes	No	Direct feed
Mansion88LiveOdds	No	Yes	Direct feed
Marathonbet	Yes	No	Scraped
MarathonBetLiveOdds	No	Yes	Scraped
Match Point	Yes	No	Direct feed
MAXbet	Yes	No	Scraped
MAXbetBU	Yes	Yes	Scraped
MAXbetLiveOdds	No	Yes	Direct feed
MAXbetSerbia	Yes	No	Direct feed
Milenium	Yes	No	Direct feed
Millennium Ba	Yes	No	Direct feed
MMMBetBU	Yes	Yes	Scraped

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection	Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
Mozzart Bet	Yes	No	Direct feed	SbbetMe	Yes	No	Scraped
myBet	Yes	No	Direct feed	SBObet	Yes	No	Scraped
Netbet	Yes	No	Direct feed	SBObet LiveOdds	No	Yes	Direct feed
NetbetIt	Yes	No	Direct feed	SBObetBU	Yes	Yes	Scraped
NGG	Yes	No	Direct feed	SCBBetBU	Yes	Yes	Scraped
Nike	Yes	No	Direct feed	Schwechat	Yes	No	Direct feed
NordicbetLiveOdds	No	Yes	Direct feed	Singapore Pools	Yes	No	Direct feed
Novibet	Yes	No	Direct feed	SingBetBU	Yes	Yes	Scraped
NSW Tab	Yes	No	Direct feed	Sky Bet	Yes	No	Scraped
Oddsen	Yes	No	Direct feed	Sky_Bets.ro	Yes	No	Direct feed
Oddset	Yes	No	Direct feed	Snai	Yes	No	Direct feed
Offside	Yes	No	Direct feed	SnaiLiveOdds	No	Yes	Scraped
Olimp	Yes	No	Direct feed	Sportingbet	Yes	No	Scraped
Opap	Yes	No	Direct feed	Sportingbet LiveOdds	No	Yes	Direct feed
Optibet	Yes	No	Direct feed	SportingBetAu	Yes	No	Scraped
Orakulas	Yes	No	Scraped	SportingIndexBetExLiveOdds	No	Yes	Direct feed
Paddy Power	Yes	No	Direct feed	Sportsbet	Yes	No	Direct feed
Paddy Power LiveOdds	No	Yes	Direct feed	SportsbetLiveOdds	No	Yes	Scraped
PaddyPowerIt	Yes	No	Direct feed	SportsSpreadLiveOdds	No	Yes	Scraped
Paf	Yes	No	Direct feed	Sporttip	Yes	No	Direct feed
Palmerbet	Yes	No	Scraped	Sportyes	Yes	No	Direct feed
Pari-Match	Yes	No	Direct feed	SpreadExBetExLiveOdds	No	Yes	Scraped
Parisport	Yes	No	Direct feed	SSBetBU	Yes	Yes	Scraped
Partypoker	Yes	No	Scraped	Stan James	Yes	No	Direct feed
Pinnacle Sports	Yes	No	Direct feed	StanJamesLiveOdds	No	Yes	Direct feed
Pinnacle Sports LiveOdds	No	Yes	Scraped	Stanleybet	Yes	No	Scraped
PinnacleBU	Yes	Yes	Scraped	Stanleybet.ro	Yes	No	Scraped
Planetwin365	Yes	No	Scraped	Star Sportwetten	Yes	No	Direct feed
Playbet	Yes	No	Scraped	Stoiximan	Yes	No	Direct feed
PMUFrance	Yes	No	Scraped	Superbast	Yes	No	Scraped
Premier Kladionica	Yes	No	Direct feed	SuperbastLiveOdds	No	Yes	Direct feed
Public Bet	Yes	No	Direct feed	SuperbetRomania	Yes	No	Direct feed
Sazka2LiveOdds	No	Yes	Direct feed	SuperSport	Yes	No	Direct feed

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
Svenska Spel	Yes	No	Direct feed
Synot Tip	Yes	No	Scraped
Tabcorp	Yes	No	Direct feed
TabNewZealand	Yes	No	Direct feed
TabNewZealandLiveOdds	No	Yes	Direct feed
TattsBet	Yes	No	Direct feed
Tempobet	Yes	No	Direct feed
TempobetLiveOdds	No	Yes	Scraped
Tipico	Yes	No	Direct feed
TipicoLiveOddsFDS	No	Yes	Direct feed
Tipos	Yes	No	Direct feed
Tipp3.at	Yes	No	Direct feed
Tippmix_TT	Yes	No	Direct feed
Tipsport	Yes	No	Direct feed
tipsport-sk.sk	Yes	No	Direct feed
Titanbet	Yes	No	Direct feed
TitanbetsLiveOdds	No	Yes	Direct feed
TomWaterHouse	Yes	No	Scraped
Tonybet	Yes	No	Scraped
Topgoal	Yes	No	Direct feed
Topgoal24ComLiveOdds	No	Yes	Direct feed
TopSport	Yes	No	Direct feed
Totesport	Yes	No	Scraped
Toto.nl	Yes	No	Direct feed
Totolotek	Yes	No	Direct feed
TotoSi	Yes	No	Direct feed
TotosiLiveOdds	No	Yes	Scraped
TrioBet	Yes	No	Direct feed
TrioBetLiveOdds	No	Yes	Direct feed
Unibet	Yes	No	Direct feed
UnibetFr	Yes	No	Direct feed
UnibetIT	Yes	No	Direct feed
UnibetLiveOdds	No	Yes	Direct feed

Bookmaker	Pre-match Odds Collected	In-play Odds Offered	Mode of Collection
Veikkaus	Yes	No	Direct feed
VeikkausFiLiveOdds	No	Yes	Direct feed
VeikkaushuoneComLiveOdds	No	Yes	Direct feed
VictoriaTip.cz	Yes	No	Direct feed
VolcanokladioniceLiveOdds	No	Yes	Direct feed
W3388BU	Yes	Yes	Scraped
Wettpunkt	Yes	No	Direct feed
William Hill	Yes	No	Direct feed
William Hill.it	Yes	No	Direct feed
WinningGoalBU	Yes	Yes	Scraped
World Of Bets	Yes	No	Direct feed
worldBet	Yes	No	Scraped
XhoiLloto	Yes	No	Direct feed
XhoiLlotoLiveOdds	No	Yes	Direct feed
Youwin	Yes	No	Direct feed

Appendix B: Odds database and screenshot details

Notes: Three matches on 15th May 2014 were missing from the Sportradar odds database: from the Danish Landspokal, odds for the match between AaB Aalborg and FC Kopenhagen in the fifth minute were missing; for the Holland Eredivisie, odds for the match between Sparta Rotterdam and Dordrecht in the third minute were missing; and for the Belgium Jupiler Pro League, odds for the match between Genk and FC Brugge in the 34th minute were missing.

Bookmaker	Date	League	Home team - Away team, Minute	Sportradar Odds	Screenshot	Market
Sbobet	07/05/14	France Ligue 1	AS Monaco- Guingamp 43'	2.11-2.4-4.76	2.17-2.5-4.60	1X2
Sbobet	07/05/14	Sweden Allsvenskan	BK Hacken - Brommapojkarna 41'	1.66-3.15-6.20	1.68-3.10-6.20	1X2
Sbobet	07/05/14	Turkey Cup	Galatasaray- Eskisehirpor 40'	1.96-2.55-5.60	2.00-2.59-5.20	1X2
Sbobet	23/08/14	France Ligue 1	Nice-Bordeaux 56'	70.00-10.00-1.04	60.00-9.75-1.042	1X2
Sbobet	23/08/14	Belgium Jupiler Pro League	Kortrijk - Oostende 54'	10.50-3.35-1.45	9.75-3.30-1.48	1X2
Sbobet	23/08/14	Spain Liga Adelante	Las Palmas-UE Llagostera 12'	1.77-3.20-4.90	1.77-3.20-5.00	1X2
Sbobet	28/08/14	Denmark 1st Div	AGF Aarhus - Fredericia 61'	1.02-11.00-95.00	1.018-11.5-100	1X2
Sbobet	28/08/14	Iceland 1st Div	HK Kopavogs - Grindavik 5'	3.10-2.95-2.17	3.10-3.00-2.15	1X2
Sbobet	28/08/14	Lithuania A league	Kruoja Pakruojis - Suduva marijampole 42'	1.5-3.05-7.80	1.50-3.05-7.80	1X2
Sbobet	05/10/14	English Premier League	Mancheher United - Everton 35'	1.17-6.60-24.00	1.19-6.00-21.00	1X2
Sbobet	05/10/14	Italy Serie A	Empoli - Palermo 49'	1.07-11.00-50.00	1.067-11.00-50.00	1X2
Sbobet	05/10/14	Germany Bundesliga 2	SV Sandhausen - FSV Frankfurt 5'	2.11-3.20-3.40	2.12-3.20-3.40	1X2
Sbobet	17/11/14	India Super League	Mumbai City - FC Goa 13'	2.38-2.84-2.88	2.40-2.74-2.96	1X2
Sbobet	17/11/14	Italy League Pro	Renate - Feralpi Salo 38'	15.00-5.60-1.14	15.00-5.60-1.14	1X2
Sbobet	17/11/14	English Conference North	Gradford park Avenue - North Ferriby United 86'	15.00-1.15-6.20	14.00-1.17-5.80	1X2
Sbobet	18/01/15	English Premier League	manchester City - Arsenal 52'	4.40-3.15-1.97	4.40-3.15-1.98	1X2
Sbobet	18/01/15	Spain La Liga	Atletico Madrid - Granada 53'	1.07-9.25-110.00	1.071-9.25-110.00	1X2
Sbobet	18/01/15	Holland Eredivisie	heracles Almelo - Excelsior SBV 67'	5.80-2.85-1.79	5.80-2.84-1.79	1X2
Sbobet	27/01/15	Turkey Cup	Mersin Idman Yurdu -Bursaspor 18'	4.50-3.30-1.81	4.50-3.25-1.81	1X2
Sbobet	27/01/15	India I- league	kalyani Bharat - Royal Wahingdoh 46'+2	1.87-2.63-4.9	1.85-2.63-5.00	1X2
Sbobet	27/01/15	Malta BOV Premiere League	Tarxien rainbows - Qormi FC 62'	5.80-2.70-1.73	5.80-2.70-1.73	1X2
Sbobet	16/02/15	Poland EkStraklasa	Gornik Zabrze - Korona Kielce 23'	2.24-2.92-3.45	2.24-2.90-3.45	1X2

Bookmaker	Date	League	Home team - Away team, Minute	Sportradar Odds	Screenshot	Market
Sbobet	16/02/15	Croatia Prva Liga	Dinamo Zagreb 66'	1.05-6.6-75.00	1.046-6.60-80.00	1X2
Sbobet	16/02/15	Turkey Super league	Genclerbirligi - Eskisehirspor 21'	1.48-3.65-7.8	1.47-3.65-7.80	1X2
Sbobet	03/03/15	International Friendly U19	Montenegro U19- Denmark U19 54'	4.6-2.15-2.39	4.70-2.08-2.46	1X2
Sbobet	03/03/15	India I- league	Mumbai FC -Bengaluru FC 8'	3.05-2.8-2.29	3.05-2.78-2.30	1X2
Sbobet	03/03/15	Russia Cup	Lokomotiv Moscow - Rubin Kazan 35'	2.20-2.43-4.7	2.20-2.35-5.00	1X2
Sbobet	12/03/15	World Cup 2018 Asia Qualifiers	Yemen - Pakistan 35'	1.1-5.4-42	1.099-5.40-42.00	1X2
Sbobet	12/03/15	UEFA Europa League	Everton - Dynamo Kyiv 34'	6.00-3.65-1.57	6.00-3.65-1.57	1X2
Sbobet	12/03/15	UEFA Europa League	Villarreal - Sevilla 73'	46.00-8.75-1.06	46.00-8.50-1.062	1X2
Sbobet	31/03/15	UEFA European U19 Championship Qualifiers	Sweden U19-Russia U19 40'	27.00-7.60-1.07	28.00-7.60-1.071	1X2
Sbobet	31/03/15	Scotland Championship	Falkirk - Cowdenbeath 2'	1.31-4.40-9.25	1.31-4.40-9.25	1X2
Sbobet	31/03/15	English Conference North	Chorley FC- Guiseley 11'	2.23-3.00-3.10	2.23-3.30-3.10	1X2
Sbobet	15/04/15	Portugal Segunda Liga	Aves- Olhanense 19'	2.29-2.72-3.65	2.29-2.72-3.65	1X2
Sbobet	15/04/15	Slovenia Cup	Maribor - NK Celje 20'	2.23-3.45-2.61	2.28-3.50-2.54	1X2
Sbobet	15/04/15	Belarus Cup	Shakhtyor Soligorsk - Dinamo Brest HT	2.24-2.26-4.30	2.24-2.26-4.3	1X2
Sbobet	05/05/15	Scotland Premiership	Inverness - Dundee United 32'	3.75-3.2-1.99	3.75-3.2-1.99	1X2
Sbobet	05/05/15	UEFA Champions League	Juventus - Real Madrid 55'	3.5-2.12-3.45	3.65-2.09-3.40	1X2
Sbobet	05/05/15	Slovakia Super League	DAC Dunajská Streda - Spartak Myjava 62'	1.17-4.30-26.00	1.17-4.30-26.00	1X2
Sbobet	11/05/15	Latvia Virsliga	Skonto Riga - BFC Daugavpils 17'	1.40-4.00-6.20	1.41-4.10-6.00	1X2
Sbobet	11/05/15	Belarus Premier League	FC Minsk - Dinamo Minsk 18'	4.70-3.35-1.63	4.70-3.35-1.63	1X2
Sbobet	11/05/15	Poland Ekstraklasa playoff	Piast Gliwice - Gornik Legnica 13'	2.26-3-3.3	2.26-2.97-3.30	1X2
Sbobet	04/03/15	Italy Lega Pro	L'Aquila Calcio - Pontedera 48'	2.53-2.05-4.2	2.54-2.03-4.30	1X2
Sbobet	04/03/15	Slovenia Prva Liga	ND Gorica - Koper 35'	7.6-3.15-1.49	8.00-3.15-1.47	1X2
Sbobet	04/03/15	Ukraine Cup	Zorya Luhansk - Dynamo Kyiv 58'	32.00-5.00-1.12	32.00-5.00-1.12	1X2
Maxbet	08/05/14	English League Championship	Brighton & Hove Albion - Derby County 55'	8.51-4.03-1.40	8.07-4.05-1.41	1X2
Maxbet	08/05/14	Iceland Premier League	Breidablik - KR Reykjavik 44'	3.58-2.76-2.28	3.59-2.74-2.29	1X2
Maxbet	08/05/14	Belgium Belgacom League	Royal Mouscron Peruwelz - Sint-Truidense 70'	3.56-1.62-4.96	3.57-1.62-4.98	1X2
Maxbet	14/06/14	Finland league	Ilves Tampere - Viikingit 20'	1.16-6.22-12.96	1.15-6.32-13.85	1X2
Maxbet	14/06/14	Thailand Premier league	Chonburi FC - Chainat FC 60'	3.99-2.69-2.01	3.85-2.65-2.07	1X2
Maxbet	14/06/14	Sweden Superettan	Degerfors - Jonkopings Sodra 79'	1.96-1.96	1.96-1.96	HC

Bookmaker	Date	League	Home team - Away team, Minute	Sportradar Odds	Screenshot	Market
Maxbet	07/07/14	Finland league	lahti - Seinajoen JK 59'	1.99-1.93	2.00-1.92	HC
Maxbet	07/07/14	Finland league	lahti - Seinajoen JK 59'	1.3-4.07-15.37	1.30-4.07-15.37	1X2
Maxbet	07/07/14	Estonia Meistriliiga	JK Sillamae Kalev - JK Tallinna Kalev 61'	1.04-7.40-42.87	1.04-7.40-42.87	1X2
Maxbet	07/07/14	Iceland Cup	BI/Bolungarvik- vikingur Reykjavik 46'	7.18-2.98-1.55	7.18-2.98-1.55	1X2
Maxbet	16/08/14	English League Championship	Leeds United - Middlesbrough 61'	4.48-1.85-3.16	4.55-1.82-3.22	1X2
Maxbet	16/08/14	English League Championship	Leeds United - Middlesbrough 61'	2.35-1.66	2.35-1.66	1X2
Maxbet	16/08/14	English premier league	Manchester united - Swansea City - 65'	2.03-2.20-9.72	2.01-2.21-10.00	1X2
Maxbet	16/08/14	France Ligue 2	Stade Brestois - Angers 53'	3.01-1.77-5.47	3.01-1.77-5.47	1X2
Maxbet	16/08/14	France Ligue 3	Stade Brestois - Angers 53'	2.26-1.71	2.26-1.71	H/C -0.25
Maxbet	03/09/14	UEFA U21 Championship 2015 Qualifiers	latvia U21 - Croatia U21 49'	24.42-6.84-1.12	24.42-6.84-1.12	1X2
Maxbet	03/09/14	Czech republic Cup	TJ Stechovice- Mlada Boleslav 6'	12.46-7.42-1.13	12.46-7.42-1.13	1X2
Maxbet	03/09/14	International Friendly	Germany - Argentina 8'	2.33-3.09-3.06	2.33-3.09-3.06	1X2
Maxbet	22/10/14	Thailand Premier league	Police United FC - Chiangrai Untied 60'	1.03-7.8-47.00	1.03-7.8-47.00	1X2
Maxbet	22/10/14	Thailand Premier league	Police United FC - Chiangrai Untied 60'	2.14-1.71	2.14-1.71	H/C -0.25
Maxbet	22/10/14	Slovenia Cup	NK Zavrc - NK Celje 22'	5.32-3.65-1.52	5.32-3.65-1.52	1X2
Maxbet	22/10/14	Slovenia Cup	NK Zavrc - NK Celje 22'	1.96-1.88	1.96-1.88	H/C -0.25
Maxbet	22/10/14	Greece Football League	kallithea FC - AOT Alimos 17'	2.03-2.73-3.83	2.03-2.71-3.87	1X2
Maxbet	22/10/14	Greece Football League	kallithea FC - AOT Alimos 17'	2.04-1.80	2.04-1.80	H/C -0.25
Maxbet	11/11/14	International Friendly	indonesia - Timor Leste 42'	1.77-2.14	1.77-2.14	H/C 0.5-1
Maxbet	11/11/14	Hungary League Cup	Szolnoki Mav FC - Diosgyor VTK 55'	3.93-2.13-2.53	3.93-2.13-2.53	1X2
Maxbet	11/11/14	Hungary League Cup	Szolnoki Mav FC - Diosgyor VTK 55'	1.79-2.05	1.75-2.09	H/C -0.25
Maxbet	11/11/14	Scotland FA Cup	Airdrieonians Fc - Greenock Morton 37'	7.13-4.54-1.39	7.13-4.54-1.39	1X2
Maxbet	11/11/14	Scotland FA Cup	Airdrieonians Fc - Greenock Morton 37'	1.88-2.04	1.89-2.03	H/C -0.25
Maxbet	19/12/14	France Ligue 2	Tours -Le Havre 80'	1.14-6.41-21.37	1.16-5.89-20.77	1X2
Maxbet	19/12/14	Holland	Jong PSV Eindhoven - Telstar 77'	4.05-1.59-4.96	4.15-1.56-5.11	1X2
Maxbet	19/12/14	English League Championship	Millwall - Bolton wanderers 83'	57.88-5.39-1.14	62.93-5.35-1.14	1X2
Maxbet	11/01/15	Holland	Panetolikos - PAS Giannina 87'	9.18-1.11-15.39	9.18-1.11-15.39	1X2

Bookmaker	Date	League	Home team - Away team, Minute	Sportradar Odds	Screenshot	Market
Maxbet	11/01/15	Spain Liga	Granada CF - Real Sociedad 62'	15.08-3.86-1.38	15.08-3.86-1.38	1X2
Maxbet	11/01/15	Italy	Napoli - Juventus 68'	4.70-1.69-4.07	4.70-1.69-4.07	1X2
Maxbet	18/02/15	Finland league	Honefoss - odd BK 10'	4.79-3.70-1.61	4.79-3.70-1.61	1X2
Maxbet	18/02/15	Portugal	SC covilha - Benfica B HT	1.03-9.00-55.00	1.03-9.00-55.00	1X2
Maxbet	18/02/15	Portugal	Leixoes - chaves 48'	3.71-2.00-3.02	3.71-2.00-3.02	1X2
Maxbet	16/02/15	International youth	Norway U19 - Portugal U19 26'	6.55-2.99-1.58	6.55-2.99-1.58	1X2
Maxbet	16/02/15	Greece Football League	Apollon Smyrnis- Kallithea FC 67'	1.18-4.19-29.52	1.18-4.19-29.52	1X2
Maxbet	16/02/15	Greece Football League	Iraklis - Aiginiakos FC 46'	1.06-6.20-53.00	1.06-6.20-53.00	1X2
Maxbet	22/02/15	Spain Liga	Tenerife - Real Valladolid HT	2.37-2.13-4.79	2.37-2.13-4.79	1X2
Maxbet	22/02/15	Greece Football League	Panathinaikos - Olympiakos 54'	1.31-4.10-13.75	1.31-4.10-13.75	1X2
Maxbet	08/02/15	Germany	karlsruher SC - Fortuna Dusseldorf 46'	1.52-2.94-9.81	1.52-2.94-9.81	1X2
Maxbet	08/02/15	English premier league	Burnley -West Bromwich Albion 74'	4.94-1.48-5.71	4.94-1.48-5.71	1X2
Maxbet	08/02/15	Italy	Ternana - Brescia 65'	12.30-3.34-1.43	11.98-3.30-1.44	1X2
Maxbet	29/01/15	Greece Football League	AEK Athens - AO Kerkyra 16'	1.61-3.45-5.29	1.61-3.44-5.31	1X2
Maxbet	29/01/15	Football international club friendly	KFUM - Strommen IF 4'	2.94-4.12-1.86	2.94-4.12-1.86	1X2
Maxbet	29/01/15	Football international club friendly	NK Osijek - NK Celje 85'	23.05-5.14-1.16	23.05-5.14-1.16	1X2
CMD368	03/03/15	India Hero I league	Mumbai FC - Bengaluru FC 38'	9.37-4.20-1.29	9.37-4.20-1.29	1X2
CMD368	03/03/15	France Cup	US Boulogne - AS saint Etienne 74'	10.18-1.48-3.33	10.18-1.48-3.33	1X2
CMD368	03/03/15	English League Championship	Bolton Wanderers FC - Reading FC 68'	1.29-4.09-16.58	1.29-4.09-16.58	1X2
CMD368	13/03/15	China Football super league	Beijing Guoan - Henan Jianye 1'	1.51-3.85-5.05	1.50-3.84-5.18	1X2
CMD368	13/03/15	China Football super league	Guangzhou Fuli - Shanghai Greenland FC 1'	2.14-3.40-2.79	2.14-3.42-2.78	1X2
CMD368	13/03/15	Poland league	Cracovia Krakow - Piast gliwice 8'	2.17-3.18-3.28	2.16-3.17-3.32	1X2
CMD368	13/03/15	Romina liga 1	FC Brasov - Universitatea Cluj 49'	3.32-1.88-3.48	3.32-1.88-3.48	1X2
CMD368	06/04/15	English League Championship	Watford FC - Middlesbrough FC 75'	1.02-10.90-127.15	1.02-10.90-127.15	1X2
CMD368	06/04/15	Belgium Jupiler pro League	Club Brugge - Standard Liege 29'	3.16-3.22-2.21	3.16-3.22-2.21	1X2
CMD368	06/04/15	Holland	Jong FC Twente - FC Emmen 31'	4.05-3.29-1.89	4.02-3.25-1.91	1X2
CMD368	06/04/15	Denmark 1st division	FC Fredericia - Viborg FF 28'	2.42-3.10-2.91	2.42-3.08-2.92	1X2
CMD368	17/04/15	Romina liga 1	CSMS Iasi - FC Brasov HT	2.27-2.24-4.21	2.27-2.24-4.21	1X2
CMD368	17/04/15	Poland league	Korona kielce - Ruch Chorzow 4'	2.24-3.10-3.20	2.24-3.10-3.20	1X2

Bookmaker	Date	League	Home team - Away team, Minute	Sportradar Odds	Screenshot	Market
CMD368	17/04/15	Ukraine 1st division	FC Ternopil - FC Zirka Kirovohrad 47'	9.78-3.61-1.35	9.78-3.61-1.35	1X2
CMD368	21/04/15	Slovenia Cup	Nk Celje - NK Maribor 73'	5.41-1.67-3.01	5.41-1.67-3.01	1X2
CMD368	29/04/15	Thailand Premier league	Tot SC - Sisaket FC HT	3.06-2.33-2.75	3.06-2.33-2.75	1X2
CMD368	29/04/15	Finland league	FF Jaro - Ilves Tampere 32'	1.31-4.73-9.50	1.29-4.87-10.05	1X2
CMD368	29/04/15	Greece Football League	Olympiacos FC - Apollon Smyrnis 30'	1.21-4.27-25.00	1.21-4.20-25.00	1X2
CMD368	29/04/15	Slovenia Cup	NK Krka - Celje 46'	56.00-8.20-1.02	56.00-8.20-1.02	1X2
CMD368	03/05/15	English premier league	Chelsea FC - Crystal Palace FC 45'	1.11-8.63-30.10	1.11-8.10-39.00	1X2
CMD368	03/05/15	Italy	Atalanta BC - SS Lazio 17'	4.91-3.24-1.86	4.42-3.22-1.88	1X2
CMD368	03/05/15	France Ligue 1	Lille OSC - RC Lens 60'	2.00-2.25-7.40	2.00-2.25-7.40	1X2
CMD368	10/05/15	Italy	SS Lazio - FC Inter Milan 15'	1.39-4.85-8.04	1.38-4.94-8.13	1X2
BWIN	12/05/15	Portugal	Varzim Guimaraes - FC Porto 69'	1.01-10.50-67.00	1.01-11.00-51.00	1X2
BWIN	12/05/15	Norway	Hønefoss Bk - Brann 50'	4.10-2.45-2.25	4.10-2.45-2.25	1X2
BWIN	12/05/15	Austria	FC Liefering - KSV 1919 36'	3.20-3.00-2.15	3.20-3.00-2.15	1X2
BWIN	12/05/15	International Club	FC Bayern - Barcelona 13'	1.60-4.10-5.50	1.50-4.33-5.75	1X2
BWIN	28/04/15	Turkey	Elazığ - Alanyaspor 46'	1.57-3.30-6.25	1.57-3.25-6.50	1X2
BWIN	28/04/15	Germany amateur	SV Rodingen - Borussia M'gladbach 24'	9.00-4.60-1.28	9.00-4.60-1.28	1X2
BWIN	28/04/15	Estonia league	FC Flora Tallinn - JK Sillamäe Kalev 46'	1.40-3.30-11.00	1.40-3.30-11.00	1X2
BWIN	28/04/15	Spain Liga	Athletic Club - Sociedad 48'	2.05-2.55-5.25	2.05-2.55-5.25	1X2
BWIN	13/03/15	China Football super league	Liaoning Whowin FC - Shandon Luneng Taishan FC 27'	7.75-3.80-1.40	7.75-3.80-1.40	1X2
BWIN	13/03/15	Ukraine	Metallurg-D Donetsk - Vorskla - D Poltava 59'	2.90-2.30-3.00	2.90-2.30-3.00	1X2
BWIN	13/03/15	Australia	UQ FC - Capalaba 72'	1.42-3.30-9.50	1.45-3.25-8.50	1X2
BWIN	30/01/15	France	Chamois Niort FC - US Creteil 13'	3.60-3.20-2.05	3.60-3.20-2.05	1X2
BWIN	30/01/15	Deutschland	Wolfsburg - FC Bayern 27'	2.85-3.40-2.35	2.85-3.40-2.35	1X2
BWIN	30/01/15	Spain Liga	Rayo Vallecano - La Coruna 17'	3.20-3.30-2.20	3.20-3.30-2.20	1X2
BWIN	19/12/14	Germany	Karlsruher SC - FSC Frankfurt 27'	1.01-14.00-51.00	1.01-14.00-51.00	1X2
BWIN	19/12/14	Italy	Mantova - AlbinoLeffe 28'	2.15-2.60-3.80	2.15-2.60-3.80	1X2
BWIN	19/12/14	France	Tours - Le Havre 69'	1.35-3.60-16.00	1.36-3.60-15.50	1X2
BWIN	06/11/14	International Club Europa	FC København - FC Brugge 14'	6.00-3.60-1.60	6.00-3.60-1.60	1X2
BWIN	06/07/14	Ireland	Athlone Town - Derry City 78'	36.00-5.50-1.12	34.00-5.50-1.13	1X2

Bookmaker	Date	League	Home team - Away team, Minute	Sportradar Odds	Screenshot	Market
BWIN	06/07/14	Sweden Superettan	VIF Brommapojkarna - Malmo FF 47'	5.25-2.60-1.90	5.25-2.65-1.87	1X2
BWIN	09/05/14	Germany amateur	FC SCHweinfurt 05- FC AUGsburg 64'	4.33-2.05-2.55	4.33-2.05-2.55	1X2
BWIN	09/05/14	Germany amateur	FC Bayern Munich 2 - FC Memmingen 66'	1.05-8.00-31.00	1.05-8.25-31.00	1X2
BWIN	09/05/14	Denmark 1st division	HB Koge - Lyngby 64'	12.50-4.00-1.30	12.50-4.00-1.30	1X2
BWIN	09/05/14	Sweden Superettan	Husqvarna FF - Osters IF 59'	12.50-3.75-1.33	11.50-3.70-1.34	1X2
BWIN	13/06/14	Iceland	Grindavik - throttur 14'	1.83-3.40-3.70	1.83-3.40-3.70	1X2
BWIN	13/06/14	Ireland	Athlone town - Drogheda 41'	1.50 - 3.50-7.00	1.50 - 3.50-7.00	1X2

Appendix C: Empirically testing the in-play model.

Testing framework

For a given starting scenario, Sportradar selected all games that matched the scenario. For example, one scenario might be the home team winning 1-0 at half-time. Let the number of games matching this i^{th} scenario be N_i . Next, we asked for how many of this subset of matches a certain event happened. For example, the event might be the home team going on to win 2-0. Let this observed number be O_i .

Sportradar provided us with the in-play model probability of the event occurring, averaged over all of the N_i games that satisfy the original scenario. Let this average probability be $\bar{p}_i$. The expected number of times, according to the model, that the event is to happen is then given by $\bar{p}_i N_i$. In this way, we can compare the number of times the event is expected to happen according to the model with the observed number of times the event happened, O_i . We now describe the statistical test in more detail.

The statistical test we use is the Chi-Square goodness-of-fit test for the multi-sample Bernoulli model. In our setup, we have N_i samples from each of i independent Bernoulli trials (each series of trials is one of the starting scenarios). Let the outcomes in the N_i matches relating to the i^{th} scenario be a vector $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,N_i})$, the elements of which equal 1 if the event in that match happened or 0 if the event did not happen. Thus, $\mathbf{X}_i$ is a random sample of size N_i from the Bernoulli distribution with unknown success parameter $p_i \in (0,1)$ for each $i \in \{1, 2, \dots, m\}$, where there are m starting scenarios.

Let the unknown parameter vector be $\mathbf{p} = (p_1, p_2, \dots, p_m)$. From the model, we have a vector of estimated probabilities $\mathbf{p}_0 = (\bar{p}_1, \bar{p}_2, \dots, \bar{p}_m) \in (0,1)^m$ and we want to test whether $H_0: \mathbf{p} = \mathbf{p}_0$, versus $H_1: \mathbf{p} \neq \mathbf{p}_0$.

For $i \in \{1, 2, \dots, m\}$ and $j \in \{0, 1\}$, let $O_{i,j}$ denote the number of times that outcome j occurs in sample $\mathbf{X}_i$. The observed frequency $O_{i,j}$ has a binomial distribution; $O_{i,1}$ has parameters N_i and p_i while $O_{i,0}$ has parameters N_i and $1-p_i$.

Now let $E_{i,0} = N_i(1-\bar{p}_i)$ be the expected number of matches, according to the model, in which the event does not happen and let $E_{i,1} = N_i\bar{p}_i$ be the expected number of matches in which the event does happen, under the null hypothesis. If N_i is large for each i , then under H_0 the following test statistic has approximately the chi-square distribution with m degrees of freedom:

$$X^2 = \sum_{i=1}^m \frac{(O_i - N_i \bar{p}_i)^2}{N_i \bar{p}_i} + \frac{(N_i - O_i - (N_i - N_i \bar{p}_i))^2}{N_i - N_i \bar{p}_i} = \sum_{i=1}^m \frac{(O_i - N_i \bar{p}_i)^2}{N_i \{\bar{p}_i(1 - \bar{p}_i)\}}$$

We note here that a modification to this chi-square test statistic is needed. This is a consequence of $\bar{p}_i$ being an average of several probabilities (and not a single probability). As such, there is less variability in the X^2 statistic above and this must be accounted for so that the test is valid. Mathematically, the modification required can be shown to be subtraction of the variance of p_i in the denominator, resulting in the following test statistic:

$$X^2 = \sum_{i=1}^m \frac{(O_i - N_i \bar{p}_i)^2}{N_i \{\bar{p}_i(1 - \bar{p}_i) - \text{var}(p_i)\}}$$

Table C.1 shows the list of events that we selected for testing. These events were chosen to represent a variety of scenarios at various stages of matches. To fully test the in-play model used by the FDS, we asked for each of the observed number of times the event occurred (O_i) and the average model estimated probability ($\bar{p}_i$) for several different leagues. The leagues were again chosen to represent all types of league and are listed in Table C.2.

In addition to testing the model on different leagues, we tested the model further, by performing these comparisons (of model estimated frequencies and observed frequencies) for each of several scenarios:

- i. a strong (relative to the away team) home team versus a weaker away team, where strong is defined as probability of winning the match of at least 50%,
- ii. weak (relative to the away team) home team versus a stronger away team, where strong is defined as probability of winning the match of at least 50%, and
- iii. all other matches.

Our reasoning for splitting the data into these groups is to test whether the model is just as sound in a range of circumstances in which it is applied.

Table C.1: Events used to test the in-play model. Note that 46 minutes refers to the first minute of the second half.

<i>Event</i>	<i>Minutes passed</i>
Home team victory	0
Drawn match	0
Away team victory	0
Match ending 2-0 when half time score is 1-0	Half time
Match ending 1-1 when current score is 0-0	46
Match ending 1-0 when half time is 0-0 and away team has received red card at some point during the first half	46
Match ending 1-0 when current score is 1-0 and home team has received red card at some point during the first half	46
Match ending 1-0 when current score is 0-0 and away team has received red card at some point during the match so far	60
Match ending 2-2 when the current score is 1-1	60
Match ending 2-2 when current score is 1-1	70
Match ending home win when current score is 0-1 after 70 mins	70
Match ending as a draw when current score is 0-1 after 80 mins	80
Away team victory when current score is 0-1 after 80 mins	80
Home team victory when current score is 1-0 after 20 mins	20

Table C.2: Leagues for which event queries were asked.

<i>League name</i>	<i>Country</i>
Albanian Superliga	Albania
Austrian Bundesliga	Austria
Cypriot 1. Division	Cyprus
French Ligue 1	France
German Bundesliga	Germany
Icelandic Úrvalsdeild	Iceland
Montenegrin First League	Montenegro
Slovakian Super Liga	Slovakia
Spanish Primera División	Spain

Test results

Using the list of starting scenarios and events (14), leagues (9), and relative team strengths (3), gives us a total of $14 \times 9 \times 3 = 378$ processes. The statistical hypothesis test we perform assumes that the observations in any one scenario are greater than 5. In cases where this was not true (e.g. there were few occasions in the Albanian Superliga when the match ended 1-0 when the half-time score was 1-0 and home team has received red card at some point during the first half), the scenarios were merged until there were more than 5. This left us with a total of $m = 263$.

For all leagues the total test statistic $X^2[263] = 272.76$ with a critical value of 301.83 (p-value = 0.327) suggesting the model fit is good. Performing the test on each league separately gave the test statistics and p-values in table C.3. In all cases the model proves to be a good fit.

Table C.3: Results of testing in-play model fit.

<i>League</i>	X^2	m	p -value
Germany	37.7439	28	0.1033
Albania	32.5924	27	0.2109
Austria	30.171	29	0.4055
Cyprus	31.2315	29	0.3546
France	27.8229	29	0.5274
Iceland	28.7836	31	0.5805
Montenegro	17.9107	29	0.9461
Slovakia	25.578	29	0.6479
Spain	40.9227	32	0.1340
All leagues	272.7567	263	0.3266

We perform two checks on the data in search of further evidence that the model fit is good. The first check is to ascertain whether the individual chi-square test statistics (calculated on each scenario) follow a chi-square distribution as theory suggests they should. We compared the histogram, the mean, variance, skewness and kurtosis of the test statistics with what would be expected from theory and they were all in agreement. The second check we did was to look at the ‘signed test statistics’. This is X^2 in the formula above, but with the modification that the numerator is not squared. Let us call this Z . Theory says that Z should follow a standard normal distribution (with mean 0 and variance 1). Again, all evidence suggested this was true and an Anderson-Darling test for normality gave a p-value of 0.261 confirming this.

To conclude, all evidence from our statistical hypothesis test suggests that the in-play model used in the FDS produces probabilities that are representative of reality.

Appendix D: Procedures for hotlisting and escalation.

