UNIVERSITY OF
BIRMINGHAM

Department of Economics

Information and Efficiency: Goal Arrival in Soccer Betting

Karen Croxson

J. James Reade

Information and Efficiency: Goal Arrival in Soccer Betting[*]

Karen Croxson[†]   J. James Reade[‡]

January 21, 2011

Abstract

In an efficient market news is incorporated into prices rapidly and completely. Attempts to test for this in financial markets have been undermined by the possibility of information leakage unobserved by the econometrician. An alternative is to switch to laboratory conditions, at the price of some artificiality. Potentially, sports betting markets offer a superior way forward: assets have terminal values and news can break remarkably cleanly, as when a goal is scored in soccer. We exploit this context to test for efficiency, applying a novel identification strategy to high-frequency data. On our evidence, prices update swiftly and fully.

JEL Classification: G14, D0, C01.
Keywords: Information, market efficiency, gambling.

1   Introduction

A matter of considerable importance in economics and finance is how relevant information becomes impounded in market prices. The significance of the topic derives partly from its theoretical pertinence: the efficient functioning of the price mechanism requires that a security's price at all times reflect its true fundamental value. It also has much to do with practical interests: traders with superior information may secure gains at the expense of the less well informed. The efficient markets hypothesis predicts that asset prices will incorporate relevant information, and in the

simplest interpretation, will do so immediately and completely.[1] Extensive efforts have been made to put matters to the test. Fama (1970) popularized the idea of considering efficiency in relation to subsets of the totality of information, focusing on three differently stringent tests. The first and most lenient test is for weak form efficiency. It requires the current price to reflect all information contained in historical prices.[2] A second test deals with semi-strong form efficiency. In a market that is semi-strong form efficient prices completely and immediately update to new information, provided that this very obviously is publicly available.[3] Finally, and most stringently, there is the notion of strong form efficiency, according to which price must at all times reflect *all* available information, even where this is held privately.[4]

Applying efficiency tests in the real world, most investigations have centered on conventional financial markets.[5] For instance, and regarding public information (the second form of test), a number of researchers have scrutinized the response of share prices to corporate events such as stock splits (Fama 1969), the release of company results (Ball 1968, Beaver 1968), merger announcements (Asquith 1983), as well as to announcements about economic variables such as the money supply (Waud 1970, Chen et al. 2003). Some investigations find support for the view that prices update efficiently but a number of others have uncovered some evidence of post-news price drift, described in the finance literature as "the tendency of individual stocks' performances following major corporate news events to persist for long periods in the same direction as the return over a short window [...]" (Jackson and Johnson 2006). For instance, in the study by Patell and Wolfson (1984) profitable trading opportunities arise following public announcements about dividend and earnings and take five to ten minutes to dissipate. Meanwhile, Chan (2003) examines returns to a subset of stocks after public news about them is released and finds evidence of post-news drift.[6]

Somewhat problematically considering the objective of such enquiries, it can be hard to tell when news actually breaks in financial markets—it is difficult to rule out information leakage not observed by the econometrician.[7] There is also the difficulty of defining normal returns. It is hard to interpret cleanly the results of tests for efficiency in financial markets as any test must assume an equilibrium model that defines normal security returns. If efficiency is rejected, this

---

[1]The efficient markets hypothesis is most commonly associated with Eugene Fama (1965; 1970; 1998). Its early origins can be traced back to the work of Louis Bachelier, who in 1900 studied the dynamics of stock price behavior (Bachelier 1900).

[2]For illustration, weak-form efficiency rules out the possibility that technical analysis techniques could be used to produce excess returns, though analysis of fundamentals still might.

[3]By implication, under semi-strong form efficiency not only technical analysis but also fundamental analysis will be powerless to deliver abnormal returns. With regard to the speed and completeness of updating, the definition of semi-strong form efficiency given in the text is the strictest interpretation. Less strict formulations exist whereby it is sufficient for efficiency that it not be possible to trade upon the relevant subset of information in such a way as to earn above-normal profits.

[4]In a market that is efficient in the strong form sense no one can earn excess returns, not even with privileged information.

[5]Vaughan-Williams (2005) offers a comprehensive review of the academic literature which has investigated information efficiency in financial markets.

[6]The literature contains ambiguous findings in terms of the sign of any inefficient reaction to news. De Bondt and Thaler (1985) find evidence that NYSE traders overreact to information due to a cognitive bias. Abarbanell and Bernard (1992) and Chan et al. (1996), meanwhile, argue that traders in financial markets adapt to new information slowly—they underreact. The literature contains several useful reviews of the evidence regarding post-event price drift (Kothari and Warner 1997, Fama 1998, and Daniel et al. 1998).

[7]Some market participants may be party to the content of announcements (or some part of this content) before these "go public" (Jarrell and Poulson 1989). See Worrell et al. (1970) for an illustrative discussion of leakage in the context of layoff announcements.

could be because the market is inefficient or because the postulated equilibrium model is incorrect. This problem—known as the joint hypothesis problem—means that market efficiency as such can never be rejected. Responding to these complications, some investigators have preferred to analyze markets in the laboratory, where conditions such as the information structure can be tightly controlled (Chamberlain 1948, and more recently, Plott and Sunder 1988, and List 2004). But while experimental settings can eliminate some concerns their artificiality raises others: what trading experience do subjects (typically students) have? Are they appropriately motivated?[8]

Potentially, sports betting markets offer a superior lens for efficiency studies, especially where news arrival is the focus. Unlike laboratory experiments, these are real markets with participants that are well motivated and often experienced. Contracts on sports outcomes (unlike equities and other financial securities) have well-defined terminal values and converge to these over a short period of time. Moreover, and most importantly, major sports news often breaks remarkably cleanly. For instance, once a soccer game has kicked off the most significant innovation in information concerns the scoring of a goal, and this event becomes common knowledge at a single identifiable point in time. This is particularly so where a game is televised, as many now are. Until very recently it was impractical to base efficiency studies around sports news; wagering was tightly controlled by traditional bookmakers (dealers), who posted prices, updated these infrequently, allowed betting only up until kick-off, and due to their business model (betting against their own customers) guarded data particularly fiercely. But from 2000, widespread Internet penetration facilitated a development in betting which would transform the industry landscape (and research possibilities) radically: the emergence of online betting exchanges.

Inspired by electronic financial exchanges, a small number of entrepreneurs began to offer web-based order-driven betting markets, enabling prospective punters to bet against each other through a live order book, with intermediation by the exchange to guarantee anonymity and remove counterparty risk. For the first time, customers could buy and sell bets at current market prices (which typically were keener prices than offered by bookmakers), submit their own limit orders, and do all this 'in-running' (as play unfolds).[9] The development proved popular with many customers and in a few short years the leading exchanges had become serious betting markets. The dominant betting exchange, Betfair (www.betfair.com), now sees trading comparable in intensity (if considerably smaller in volume) to activity on the world's leading financial exchanges. Betfair was one of the very first betting exchanges onto the market and has grown to be overwhelmingly the largest; its turnover of over \$50m per week accounts for 90% of all exchange-based betting activity worldwide and it currently has over two million registered users. Two million trades a day—six times the number of trades on the London Stock Exchange—are processed through Betfair markets, and the exchange now covers a vast variety of events, mostly sporting.[10]

From an academic perspective, the success of Betfair presents an attractive research opportunity. We exploit data extracted from the exchange at high frequency to conduct a novel and remarkably clean test for semi-strong form efficiency. The data comprise second-by-second snapshots of Betfair's live order book for professional soccer games, featuring in-running prices and volumes related to betting on the outcomes of 1,206 matches. Included in the sample are recent

---

[8]See Levitt and List (2007) for a recent consideration of factors affecting the generalizability of laboratory findings, including the extent to and manner in which subjects are scrutinized in their decision-making.

[9]A limit order is a speculative order to place a bet at a price not worse than some specified 'limit' price and for a stated volume. The price specified is more attractive than the current market price and the order will be 'filled' only if the market moves in a favorable direction.

[10]"Betfair Makes Online Odds on AC Milan, Hillary Clinton, Weather," *Bloomberg News*, September 6, 2005. Article available online at: http://quote.bloomberg.com/apps/news?pid=nifea&&sid=a8Y11XQeIcyY

English Premiership matches (547), games played as part of the Euro 2008 Championships (101), games from the Champions League (165), the Scottish Premiership league (64), the UEFA Cup (249), the Intertoto Cup (14), the Asian Cup (24), and a number of international friendlies (42). The average match is heavily traded, with over $6m bet in total, and half of this in-running.

The major news in a soccer match concerns the arrival of goals. Goals arrive infrequently and tend to be material to match outcomes. If the betting markets in our sample are semi-strong form efficient then prices should respond immediately and completely to such news. The paper implements a complementary set of tests to investigate.

Immediacy of reaction implies a jump in price, and it is straightforward to test for this. We carry out disaggregated analysis on our full sample and confirm that prices jump when a goal is scored; on average the scorer's win probability jumps up by 22 points. Drilling into the sample, the size of initial jump can be considerably smaller or larger than this depending intuitively on such factors as the precise stage of play (later goals tend to have a greater impact) and the realized scoreline (goals that change the default match outcome have the biggest effect on prices).

To ascertain whether an initial jump represents complete updating or merely the start of a more sluggish updating process is somewhat less straightforward. Consider that, once a soccer match is underway, participants should update to major news such as a goal, but also to the continual flow of more minor news inherent in the ticking down of the playing clock, the arrival of cards for offences, free-kicks, and injuries. As playing time elapses efficient betting prices should drift continually, therefore, reaching their terminal values (1 or 0 in probability terms) by the end of the game. By implication, properly identifying inefficient drift in a betting price during minutes of play would require positing some model of 'efficient' drift. But quite what efficient drift looks like will be match specific, depending on such factors as the teams competing, the new scoreline, and time left on the clock. We implement three complementary approaches to testing for efficiency in this setting.

Our first approach introduces an identification strategy which allows us to study the incorporation of major news in individual contracts whilst side-stepping the potential complication of 'efficient' drift. Our strategy involves exploiting the (virtually) newsless window provided by the half-time interval in play. Soccer games feature two periods of play—a first half and a second half, each of 45 minutes plus a few minutes of 'injury time' (added on by the referee to compensate for stoppages). Between these periods the match stops completely for 15 minutes, but betting related to the match outcome continues apace. We exploit this window as an opportunity to identify potential inefficient goal-related drift simply and cleanly. Concretely, we study the reaction to goals that arrive on the cusp of half time. Our sample features 160 goals that arrive within five minutes of the precise end of the first half. Looking more closely at these 'cusp goals', 53 are scored in the final minute of first half play. These goals provide the basis for a particularly strong test for semi-strong form efficiency. Focusing on the reaction of half-time prices to these goals, we implement both a test for statistical efficiency, using regression methodology, and a test for economic efficiency. Our test for economic efficiency asks whether a hypothetical trader could make money during the half-time interval by exploiting goal-related price drift during the break. We are unable to reject the efficiency hypothesis that a cusp goal immediately shifts price but does not cause this to drift during the interval: prices update so swiftly and completely that the news of a goal is fully digested by the time the break commences, even where the goal occurs just moments before the end of play.

The key strengths of this test are simplicity and cleanness. A potential weakness relates to the potential for any efficiency finding to be specific to the half-time interval. One might suspect, for instance, that our inability to detect sluggishness in updating over the break could be due to

a lull in trading during this time. We deal with this concern by tracking and reporting half-time trading activity; betting interest remains healthy during the break, and certainly half-time trading is strong in games which feature cusp goals.[11] One might worry that the half-time finding does not generalize to minutes of play for other reasons: different trader types could be active during half time or goals might interact with the arrival of more minor news when the match is in progress.

For robustness, therefore, and capitalizing on the richness of our data set, we deploy two further testing strategies. Both are intended to provide a more direct look at the incorporation of news during minutes of play. The first exploits our large sample size in exploring whether *average* in-play prices display inefficient drift. An informationally efficient asset price should satisfy the martingale property—the current price should be the market's best guess of the price next period. Yet, as noted already, individual in-play contracts will display almost continuous drift, regardless of the efficiency of the underlying contracts. Consequently, testing for a martingale in the price series for an individual contract, or a small set of contracts, cannot yield meaningful conclusions. The particular importance of sample size in this context can be understood by recognizing that a test for absence of drift in in-play prices is a close relation to the standard calibration tests now familiar to readers of the prediction markets literature (for instance, Wolfers and Zitzewitz 2004 and references therein). Like most prediction markets contracts (for example, 'Obama to Win'), the Betfair 'match outcome' contracts that we study have a binary payoff structure—they pay out 1 in the event of a win, and 0 otherwise. This means that it is never possible to tell whether an individual binary contract priced at say, 0.6, was priced efficiently. It is, however, possible to consider a sample of such contracts priced at, say, 0.6, and *provided the sample is large enough* test meaningfully for efficiency by asking whether these 'win' (have a final price of 1) 60% of the time. This is a standard calibration test. It is equivalent to asking whether the average final price is 0.6—or whether *on average* the contract prices do not drift between now and the end of the event. Hence, such a test can be considered a special case of a martingale test where the future price considered happens to be the final price.

Typically those conducting standard calibration tests plot contract prices (in probability terms) against 'win' frequency, and carry out standard diagnostic tests to assess closeness of fit to the 45–degree line. In an efficient market all the data points should lie on the 45–degree line. But if the sample of contracts is too small then the fit to the 45–degree line will be poor, regardless of efficiency. In the extreme, imagine a calibration test conducted on a sample that includes only ten contracts, one in each of ten price intervals: (0–0.1), [0.1–0.2) . . . , [0.8–0.9), [0.9–1). Each of these ten contracts must go on to win or lose. With a single contract in the interval [0.4–0.5), either 100% or 0% (but never 40–50 %) of the contracts in this bracket will 'win'. Thus, for a small sample like this, it is very easy to see how the results (evidence for efficiency) will have little meaning.[12] Similarly, tests for average post-goal drift will have little bite unless the sample is sufficiently large, providing good coverage of the various price points. With a large enough sample we should find that for those contracts priced efficiently at 0.9 immediately following a goal the

---

[11]Weak half-time trading is observed in a few exceptional cases. For instance in the Premiership match between Wigan and Liverpool, the latter had built up a virtually unassailable 4–0 lead by half time. This early domination appears to have killed interest in the match odds market.

[12]See, for instance, Christiansen (2007) for a practical illustration of the problems of conducting standard calibration tests on a small sample of prediction markets. His sample featured 39 markets trading contracts related to the outcomes of rowing contests. Standard calibration on these data, using intervals of ten percentage points, shows a relatively poor calibration curve. The poor calibration is due to the small number of data points in each interval. For example, only one contract in his sample had a price (probability) of over 80% (a contract on a previously unbeaten Great Britain boat to win a Henley Regatta) but that contract lost, undermining the overall calibration (see his Figure 2).

average price in the next minute or so is 0.9. Indeed the average price should remain at 0.9 for the remainder of the game. We utilize the large size of our sample to implement such a test in the paper, focusing on the stability of the average price series in the fifteen minutes following goals. We are unable to find signs of significant in-play drift in average prices following goals and so fail to reject the hypothesis that these markets update efficiently during minutes of play.

Our second approach to testing efficiency in-play is more elaborate; it involves modelling in-running match outcome probabilities in order to estimate 'efficient' time-related drift for each match. Concretely, we begin by 'reverse engineering' the goal arrival process from historical match data. The bivariate Poisson model suggested in Karlis and Ntzoufras (2003) provides a natural starting point for this exercise. Following this, we combine the fitted goal arrival model with a multinomial model to back out 'efficient' in-running match outcome probabilities for the games in our sample. The final step is to compare the two series—our 'efficient' price series and actual observed Betfair prices—with a view to drawing inferences regarding efficiency. We find that the probabilities implied by Betfair prices closely track those implied by the selected Poisson process, which suggests that most of the in-play drift we observe can be ascribed to efficient updating to the clock. There are some differences between the two series and we discuss these further in the paper. This form of efficiency analysis is notably less clean than our half-time identification: the joint hypothesis problem is severe as our test relies on the ability of the posited Poisson model (which cannot incorporate in-play developments unobservable to the econometrician) to describe 'efficient' in-play returns. Nevertheless, the modelling exercise offers an insightful and again more direct perspective on efficiency during minutes of play.

In addition to previous work on efficiency in financial markets, our study is related to a collection of earlier studies of efficiency in sports betting markets. The vast majority of previous betting analyses are based on low frequency prices (typically bookmaker odds) sampled prior to the start of a live event (betting in-running being a very recent development). For instance, studies by Golec and Tamarkin (1991), and Gray and Gray (1997) both examine efficiency in the NFL betting markets using the closing spreads of Las Vegas sportsbooks. Closing spreads are the final prices quoted by the bookmakers shortly before game time. Vaughan-Williams 2005 provides a recent and thorough review of previous work on betting and efficiency. In its exploitation of in-running betting exchange data, and its focus on goal arrival in soccer betting, our work is most closely related to a recent paper by Gil and Levitt (2007). Gil and Levitt analyze data from the Intrade exchange (www.intrade.com), which until recently operated markets for sports-related bets.[13] Considering fifty matches from the 2002 Soccer World Cup, the authors implement an event-study methodology to look at updating to goals during minutes of play. They report that Intrade prices, though they respond strongly to a team scoring, trend for ten to fifteen minutes after the goal is registered. On the face of it, this drift appears reminiscent of the post-news drift found in some financial market studies. Gil and Levitt (2007) interpret the drift they observe in Intrade's markets as evidence of informational inefficiency—prices, they suggest, update sluggishly to the news of a goal. The analysis in their paper is insightful in many ways (for one, the data come disaggregated at the individual trader level, and the authors are able to document the endogenous emergence of market makers), but as a test for semi-strong form efficiency it suffers severe limitations. The main concern relates to data quality: Intrade soccer markets attract very few traders (on average just 75 per game); these people make very infrequent trades (an average game attracts 100–200 trades and

---

[13]To avoid difficulties with US law, which considers wagering on sports to be gambling, Intrade-Tradesports, a single operator at the time of the 2002 World Cup, has since split into two separately registered companies with different activities: Tradesports deals with sports betting, whereas Intrade now operates as a prediction market focused exclusively on non-sports events.

features several minutes in which trades do not occur); and betting volumes typically are low (just $1.5m is traded in total across the full set of fifty matches). The first of our own in-play efficiency tests, which applies a comparable regression methodology to a large sample of Betfair markets, fails to reject the efficient market hypothesis. We can conceive of a few possible explanations for the apparent discrepancy in our respective findings. It could be that the Intrade markets Gil and Levitt (2007) study are inefficient, perhaps because of cognitive biases on the part of traders or simply because of the markets are very thin. An alternative explanation is that this particular test for efficiency—being a test for whether there is drift in average in-running prices—is not robust to their relatively small sample size, for the reasons discussed above.

We conclude our study by using the dataset to take a closer look at whether and how liquidity affects a market's ability to aggregate information. This issue stands at the heart of an important and unresolved debate in the finance literature. It has considerable academic but also practical relevance, perhaps most obviously in the field of prediction markets design.[14] A number of theoretical articles establish that liquidity is related to the informational efficiency of markets but the sign of the relationship is disputed. In the first view, increased liquidity lowers the transaction cost (price impact) for informed arbitrageurs and creates greater incentives to acquire information, leading to improved informational efficiency. In the second view, liquidity is a proxy for non-informational trading (noise trading), which may harm informational efficiency (De Long et al. 1990). A further possibility is that liquidity and informational efficiency are unrelated. Whilst some empirical studies have lent support to the first view, that securities mispricing is greater in illiquid markets (Kumar and Lee 2006; Sadka and Scherbina 2007; and Chordia, Roll, and Subrahmanyam 2008), some other investigations appear to support the idea that liquidity worsens mispricing (Tetlock 2007, Hartzmark and Solomon 2008). Tetlock's (2007) empirical analysis utilizes asset prices sampled at thirty minute intervals from the online prediction market TradeSports. As he points out in his paper, the advantage of working with such a data set is that the assets traded have short lives and so reach their terminal values quickly. He finds "strong empirical support for the hypothesis that the prices of illiquid securities converge more quickly toward their terminal cash flows," so supporting the second strand of theoretical work. A limitation of the study, from the perspective of semi-strong form efficiency, is that the author does not directly observe event-relevant information.

Across our dataset as a whole, we observe considerable variation in liquidity per game, with betting volume of over $50m in the most heavily traded match, compared with just over $0.05m in the least traded. We exploit this variation, together with the clean and observable arrival of event-relevant information, to offer a new perspective on the link between liquidity and efficiency. Our preliminary investigations would appear to support the view that liquidity and informational efficiency are not strongly related in these markets, at least over the liquidity range we observe. Liquidity can be defined in a number of ways, and in the current analysis, we use in-play trading volume as our proxy. Future work could deploy more sophisticated liquidity measures. It would be of further interest to investigate the effect of news arrival on intra-match liquidity.

The rest of our paper is organized as follows. The next section provides further background on

---

[14]Recent years have seen a growth in real-world interest in the topic of prediction markets, and considerable hype about the potential for markets to revolutionize forecasting and decision making. Yet the formal literature on prediction markets is very underdeveloped and has yet to investigate with sufficient rigor whether the information such markets generate can be relied on for decision making. Whilst evidence from some existing studies (many of the Iowa Election markets) is encouraging, and several researchers have recently emphasized the ability of markets to improve decisions (Hanson 2002, Hahn and Tetlock 2005, and Sunstein 2006, among others), considerably more work is needed. For further information the reader is referred to the recent survey article by Wolfers and Zitzewitz 2004 and to Hahn and Tetlock Eds., an excellent compendium of papers on information markets.

the betting industry and discusses Betfair in more detail, whilst Section 3 describes the data set used in this study. Section 4 discusses estimation strategy and presents the main findings regarding market efficiency. Section 5 briefly investigates the possible link between informational efficiency and liquidity. Concluding remarks are set out in Section 6. Section 7 is an Appendix containing supplementary materials referred to in the main text.

2   Betting and the Betfair Exchange

Traditionally, betting markets have been run by a closed community of licensed dealers, known as bookmakers. Bookmakers are similar to market makers in financial markets; they establish and maintain liquid markets by quoting prices at which they will deal. In betting, the prices are termed "odds" and the most common type of bet is known as a fixed-odds bet. Suppose party A wishes to back (bet on) some outcome and party B wishes to lay (bet against) the same. Then under a fixed-odds bet, A agrees to pay B a certain amount (the backer's "stake") if the outcome fails to materialize, and B agrees to pay A the same stake multiplied by pre-agreed (hence "fixed") odds if instead it does. For example, A might stake $100 at odds of 3 : 1 ('three to one') that Argentina will win the World Cup. In this case, she collects $300 from B if Argentina succeed, but otherwise B keeps her $100 stake. When betting with bookmakers, customers are restricted to backing outcomes only; the bookmaker plays the role of party B, taking the lay side to every bet.

Odds relate inversely to the probabilities associated with particular outcomes.[15] For instance, odds of 3 : 1 imply a view that Argentina is three times more likely to fail than to succeed (a 25% probability of Argentine victory).[16] Bookmakers rely on in-house gambling experts to assess the likelihood of different outcomes and to compile a set of odds accordingly. As the event draws closer the odds can be adjusted, reflecting the arrival of relevant information and the bookmaker's desire to maintain a balanced book. Odds are described as "fair" when the implied probabilities sum to one, but built into the set of prices offered by the bookmaker is a return for liquidity services (known in betting circles as the "overround" or "vigorish") such that the sum of probabilities exceeds one. In 1999, this bookmaking model was still the only model of betting, and bookmakers belonged to an exclusive and profitable club. In the UK, one of the world's key betting markets, it was illegal for anyone other than a licensed bookmaker to accept bets and a handful of major players (William Hill, Ladbrokes, Coral) dominated the market. The overround stood at a healthy twenty two per cent.[17]

The arrival in 2000 of online betting exchanges marked a revolution in the industry. The leading exchanges are essentially order-driven markets in fixed-odds bets, allowing individual punters to bet with each other directly, thereby disintermediating the bookmaker. This means that exchange bettors can and do lay individual outcomes, contrary to the standard bookmaking model. In addition, exchanges allow customers to place bets "in-running," once an event is underway. This is felt to have created a significantly more exciting betting experience. Typically, customers are charged a small commission for exchange betting services, but the exchange does not otherwise impose any overround. Compared with bookmakers' odds, exchange prices, at least for popular

---

[15]The interpretation of betting prices as probabilities is a somewhat debated area. The interested reader is referred to Wolfers and Zitzewitz (2006) and the articles cited therein, particularly Manski (2006).

[16]In this example, the odds are quoted in so-called fractional form. An alternative is to quote decimal odds, in which case the stake is included in the quoted multiple, so that 3 : 1 becomes 4. This is convenient because the implied probability is then obtained simply by inverting the decimal odds and normalizing.

[17]Merrill Lynch Research, 17 January 2006.

events, have tended to be highly competitive.[18] The real hurdle for exchanges has been to achieve sufficient liquidity. Betfair was one of the first exchanges to market, and is now by far the largest. It levies a standard commission of 5% on winning bets, falling to 2% for the heaviest users.[19] Betfair's early entry into the market and its decision to run with a model much closer to a standard financial exchange than some of its competitors (notably Flutter.com) are thought to have been pivotal its success.[20] Volumes on the exchange are estimated to have doubled from $5.23bn to $11.06bn between 2003 and 2004, and almost doubled again between 2004 and 2005.[21] These growth rates are well ahead of those for the gambling market generally. Figure 1 benchmarks Betfair to the world's largest financial exchanges in terms of trade frequency.[22]

Betfair processes around two million trades a day—six times the number of trades on the London Stock Exchange. And in the past few years the search term 'betfair' has overtaken 'FTSE' in popularity on google.co.uk (Figure 2).

The selection of markets Betfair offers is vast and covers most sporting events of popular interest, together with many non-sporting events (such as key political events and reality TV). Horse racing dominates exchange turnover, followed by soccer. Within soccer betting, customers can place bets related to the 'Outright Winner' of a particular league or tournament, or the 'Top Scorer' of the competition, for instance. Meanwhile, 'Match Odds' markets allow betting on the outcome of individual games, by backing (betting on) or laying (betting against) the 'Home Win', 'Away Win', or 'Draw'. For those with less conventional betting preferences, there are markets such as 'Over 2.5 goals,' 'Half-Time Score' and exotic bets, such as Asian Handicaps and Multipliers.

Suppose a user wished to bet on the outcome of a recent Premiership encounter between Arsenal and Manchester United. Figure 3 shows the order book shortly before kick off.[23]

The order book indicates, among other things, that should she wish to back Manchester United, she might immediately stake up to $16,784 at odds of 3.3, and up to a further $53,140 at slightly

---

[18]The dominant exchange, Betfair, claims its prices are on average 20% more generous than bookmakers'. Ozgit (2005) confirms the competitiveness of Betfair's pricing in the Basketball markets, although he does draw attention to the failure of exchange sometimes to offer sufficient liquidity at inside (best) prices. He notes that punters wishing to place large bets can be better off taking their custom back to bookmakers rather than "walking down the order book" at the exchange, accepting increasingly unattractive prices to get large orders filled.

[19]On the 8th of September 2008 Betfair announced the introduction from 22nd September of an extra "Premium Charge" to be paid by those customers whose winnings over the preceding 60 weeks have reached a certain threshold. Such gamblers will be required to pay 20% of their profits to Betfair in commission or other charges. The stipulated winnings threshold was set so high that the vast majority of Betfair users have so far been entirely unaffected by this innovation to the charging scheme. Nevertheless the development has proved controversial: http://www.guardian.co.uk/sport/2008/sep/16/horseracing

[20]Flutter.com, founded in February 1999 by American management consultants, was the first person-to-person betting site. Des Laffey (2005) analyzes some of the operational and marketing differences likely to have led to Betfair's dominance over (and eventual merger with) its main rival, despite the Flutter managing to attract a comparatively huge amount of financial backing for its launch: "Flutter believed that they could thrive by facilitating social bets between friends, for example about who would win a game of golf, and also limited the value and frequency of bets allowed." "Flutter's website was not based around the Betfair idea of matching pools of money from backers and layers, instead requiring a complete match between a single backer and a single layer. Multiple transactions on an event by a punter on Flutter were also treated separately which led to inefficiency whilst the Betfair model recognised mutually exclusive outcomes."

[21]Merrill Lynch Research, 17th January 2006.

[22]This figure is based on analysis undertaken by Stephen Roman, Analyst, FXCM, New York, commissioned and reported by prediction markets blog midas.org.

[23]The standard view of the Betfair order book shows the best three prices (and corresponding available volumes) on the back and the lay side. By clicking on the team name it is possible to view the full order book showing any prices and volumes available beyond the first three steps of the book, along with historical prices charts for each selection.
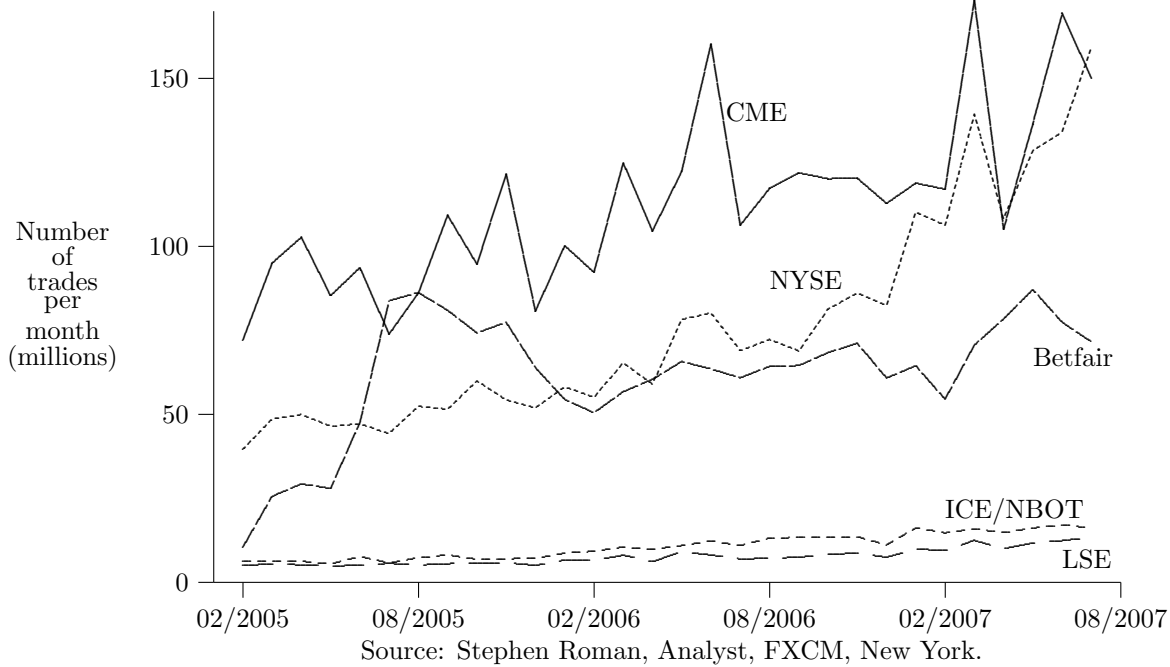
Figure 1: Betfair vs. major financial exchanges: daily trading intensity.

less attractive odds of 3.25. Betfair uses decimal odds which are inclusive of stake. So a $10 bet to back Arsenal at odds of 2.58 would result in a gross return of $25.80 ($15.80 profit plus $10 stake). All odds are displayed from the backer's point of view. Thus, 2.6 and $16,289 on the 'Lay' side of that market implies that someone (or some combination of users) has submitted limit-orders hoping to back Arsenal asking for odds of 2.6 (i.e., slight better than the prevailing market odds). If she were to accept $10 of this, by placing a lay order at 2.6, the user would be betting against Arsenal and risking $26 to win $10.

3 Data

The data deployed in this paper comprise second-by-second prices and volumes from Betfair's 'Match Odds' markets for 1,206 professional soccer games. We capture the evolution of each time series in-running (as the match is being played). Included in the sample are recent English Premiership matches (547), games played as part of the Euro 2008 Championships (101), games from the Champions League (165), the Scottish Premiership league (64), the UEFA Cup (249), the Intertoto Cup (14), the Asian Cup (24), and a number of international friendlies (42).[24] As discussed in the previous section, 'Match Odds' soccer markets offer betting on the basic outcome of each match (Home Win, Away Win and Draw). Explicitly, the data set contains the following variables for each match:

1. timestamp;

2. the game outcome to which the order book relates (e.g. Home Win);

---

[24]Due to technical and practical limitations on the part of the first author (who acquired the data set) it was not possible to collect data for all recent matches in these competitions. The sample selection is random, however. The Appendix contains summary information for matches sampled.
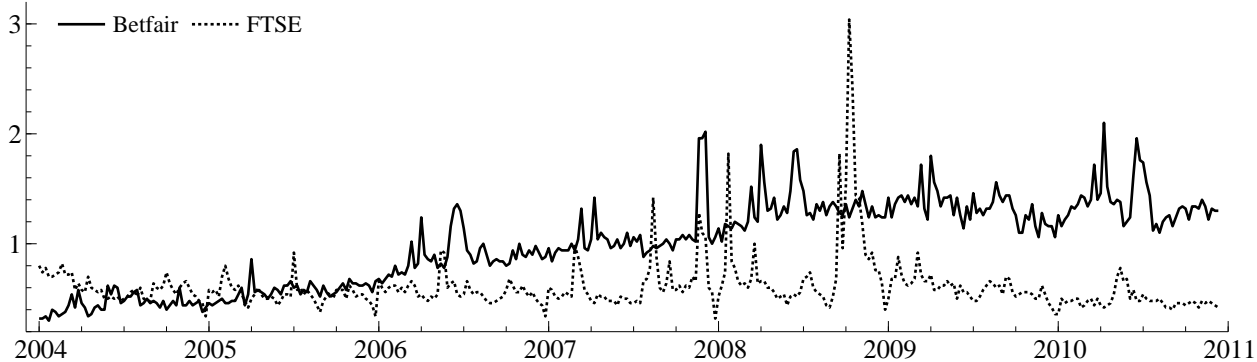
10

Figure 2: Frequency of google searches for "betfair" vs. "FTSE". Scaling is relative to average searches for Betfair such that a value of one means that at that point the number of searches for either Betfair or FTSE was equal to the average number of searches over the entire sample period.

Figure 3: Betfair order book: Arsenal vs. Manchester United

3. the best three prices to "back" this outcome and the volumes available to bet at each price;

4. the best three prices to "lay" this outcome and the volumes available at each price;

5. whether the market is "in-play;"

6. whether the market is "suspended;"

7. the total cumulative volume traded on the Match Odds market for this game.

The market for a particular match is "in-play" when that match is in progress. As mentioned in the previous section, Match Odds markets for professional soccer matches tend to be heavily traded, particularly during the games themselves. Across the sample as a whole, the average match is heavily traded, with over $6m staked per game Typically, half of this is bet in-running, which equates to $31,627 traded per minute and $527 per second. Behind this headline average, the betting interest is quite variable across matches, with betting volume of $50m in the most heavily traded match, compared with just over $0.05m in the least traded. Summary statistics for the

11

1,206 matches in our sample are reported in the Appendix. Many English Premiership games are now televised (in the case of English Premiership matches, either on Sky Sports or Premiership Plus subscription channels) and the sample features an interesting mix of televised and untelevised encounters. Television coverage tends to boost associated Betfair trading significantly.

Betfair briefly suspends its in-play soccer markets at kick-off and then briefly again upon the occurrence of what it defines to by a "Material Event." In the context of soccer, a Material Event is the scoring of a goal, the award of a penalty, or the sending off of a player (the awarding of a red card). As far as the Match Odds for a game are concerned, the scoring of a goal is the most important piece of news. Goals arrive fairly infrequently; across our sample there are on average 2.55 goals per match. During a goal-related trading suspension Betfair discards any unfilled orders, thereby clearing out the entire betting order book. When the order book reopens the odds have shifted, reflecting updating by the market about the relative chances of the Home Win, Away Win, and Draw. Figure 4 illustrates the suspension of trading and subsequent price updating in a recent English Premiership encounter between Bolton and Tottenham Hotspur.

(Kick-off 17:15 on 09/23/2006, televised, Bolton wins 2-0 with goals at 17:25 and 17:29)
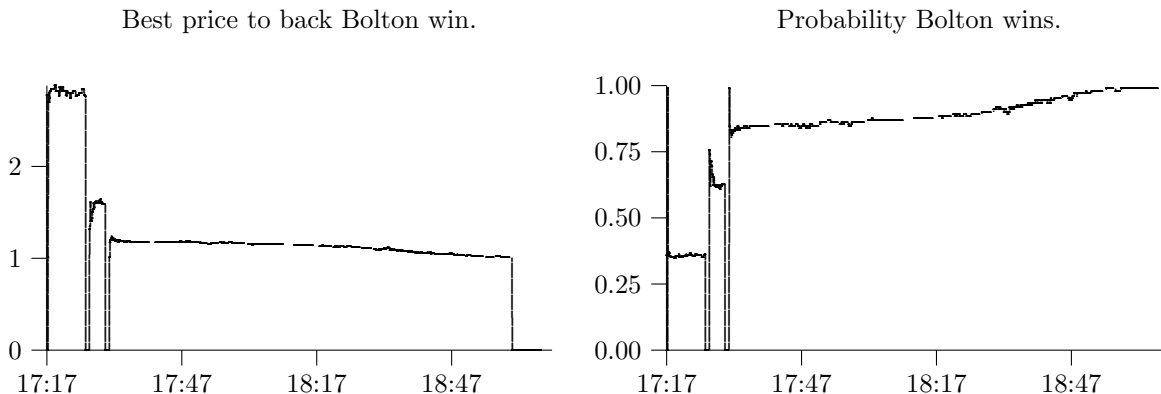
Figure 4: Price evolution in-play: Bolton Wanderers vs. Tottenham Hotspur.

The left hand panel plots the "in-running" time series for the best price to back Bolton to win, and this price is converted into an implied probability in the right hand panel.[25] Bolton scores two goals in quick succession, at 17:25 and 17:29, and the vertical lines clearly visible at both points in time represent the associated suspension of trading and removal of all unfilled orders. Once the market reopens following a goal, the book fills up quickly with new orders at new prices. This updating manifests in clear jumps in price and implied probability.

Data concerning the exact timing of goals and any other Material Events were obtained from SportingLife.com. Sometimes matches kick off slightly later than officially planned (it is not atypical for a game to begin at 15:03 rather than 15:00, for example) and at the end of each half the referee adds on a small amount of extra playing time (usually called "injury time"). This added time is to compensate for the loss of standard playing time due to injuries and other stoppages. Betfair records the precise duration of each half for the purposes of officially closing and settling the many markets it offers in relation to each soccer game. This allows us to infer exact match timings for the games in our sample, to the nearest second.

---

[25]Implied probabilities are computed as 1/(decimal odds). For instance, decimal odds of 4 would a 25% probability. (As actual decimal odds often sum to more than 1, a normalization is applied to ensure that implied probabilities sum to unity.)

4   Testing for Semi-Strong Form Efficiency

If Betfair markets are semi-strong form efficient, prices (and the probabilities these imply) should update to public news rapidly and fully. In this section, we assess the immediacy and completeness of the Betfair price response to goals scored during the soccer games in our sample.

4.1   Do Markets Respond Immediately to Goals?

It is straightforward to confirm that prices respond immediately to the news of a goal. We have preliminary evidence of this already: from plots such as those in Figure 4 it is apparent that the price level has jumped between the market closing upon a goal being scored and it reopening shortly afterwards.[26] Table 1 looks more closely at the magnitude of this immediate market reaction to the 2,528 goals in our sample.[27]

		Total	(1)	(2) 0	(3) 1 (+)	(4) 1 (−)	(5) 2 (+)	(6) 2 (−)	(7) #≥3
				Goal Difference Immediately Following Goal					
(1)	All goals	2528	**0.22**	0.16	0.30	0.06	0.12	0.03	0.01
(2)	Home goals	1491	0.21	0.18	0.30	0.08	0.12	0.07	0.01
(3)	Away goals	1037	0.23	0.15	0.33	0.04	0.14	0.01	N/A
(4)	Goals by favorites	1126	0.21	0.20	0.30	0.09	0.11	0.04	0.01
(5)	Goals by outsiders	926	0.23	0.13	0.34	0.03	0.17	0.02	0.03
(7)	International	352	0.21	0.18	0.29	0.06	0.13	0.06	0.01
(8)	Domestic	2176	0.22	0.16	0.30	0.06	0.12	0.02	0.01
(6)	Late goals (≥ 80mins)	357	0.30	0.12	0.64	0.02	0.05	0.06	N/A

Table 1: Immediate goal-induced change in scorer's win probability.

The value in bold in the top left cell of this table indicates that a goal on average induces an immediate 22 point increase in the scoring team's win probability. This statistic is in line with findings reported elsewhere regarding goal impact. For instance, in Gil and Levitt (2007) a World Cup 2002 goal induces a change of between twenty and thirty points in the implied probability that the scoring team wins. Intuitively, there is considerable variation behind this simple average, as the other cells in Table 1 suggest. Rows (2)–(8) consider subsets of goals grouped according to scorer type (whether the scoring side is the home team or the away team, the a priori favourite or the underdog), lateness in the game, and whether or not the match considered is an international fixture. Columns (2)–(7) classify goals according to the goal difference they create. Goals that change the status quo outcome (e.g. from a draw to a win) have the greatest impact, and among these, those that put the scoring side ahead have a greater impact (column 3) than those that bring the scoring team level (column 2). Goals that merely extend a side's lead tend to have smaller effects. For example, a goal that increases a team's lead to two goals (column 5) increases its win probability by just 12 points on average. Goals scored in international matches in our sample tend to have a slightly smaller impact on the scoring side's win probability, raising this by

---

[26]Recall from the previous discussion that Betfair suspends the market briefly in the event of a goal.

[27]In computing this shift, we look at the market price immediately before the goal is registered and compare this to the price shortly after the goal-induced trading suspension, once reasonable liquidity has returned to market. We exclude from the full sample any goals for which we have insufficient data around the time of the goal.

21 probability points on average (row 6, column 1).[28] Outcome-changing goals that occur towards the end of the game tend to have the greatest impact, as would be expected. A goal scored late in the game—after the 80th minute—and which puts a team in the lead, adds on average 64 points to the probability that they go on to win the game (row 8, column 3).

4.2   Is the Immediate Response also Complete?

4.2.1   A Complication: Efficient Prices *Must* Drift Continually During Play

Whereas we can be comfortable on the strength of this evidence that there is an immediate reaction to goals in these markets, it is somewhat more complicated to ascertain whether the jumps observed reflect complete Bayesian updating or simply mark the beginning of an updating process that takes some time to complete. To pursue this question of completeness, we might think about comparing the new post-jump price level with the level several minutes later (assuming no further goals in that time), and construing any significant difference between these as evidence of informational inefficiency. However, to identify inefficiency in sports betting markets that are in-play one confronts an interesting and non-trivial complication: some amount of price drifting is perfectly consistent with, and indeed evidence for, market efficiency since rational participants would be expected continually to update to minor news during the game, not least the passage of playing time without a goal. Consequently, if informational (in)efficiency is to be inferred from the post-goal evolution of prices, a strategy must be found to separate out 'efficient' drift (rational updating to the ticking down of the clock and other minor in-match news) from possible drift due to sluggish incorporation of major news (evidence for semi-strong form inefficiency).

To appreciate the nature and inevitability of time-related drift in in-play prices, consider that as match time elapses the likelihood of further goals fades, and the probability thus grows that the current standing of the teams will come to reflect the final match outcome. In the absence of further goals, the probability associated with the contract that would win under the status quo scoreline should drift upwards over time, reaching one by the end of play. By contrast, the probabilities associated with the other possible outcomes of the game should drift downwards towards zero. Figure 5 illustrates time-related drift using the prices associated with two Premiership fixtures.
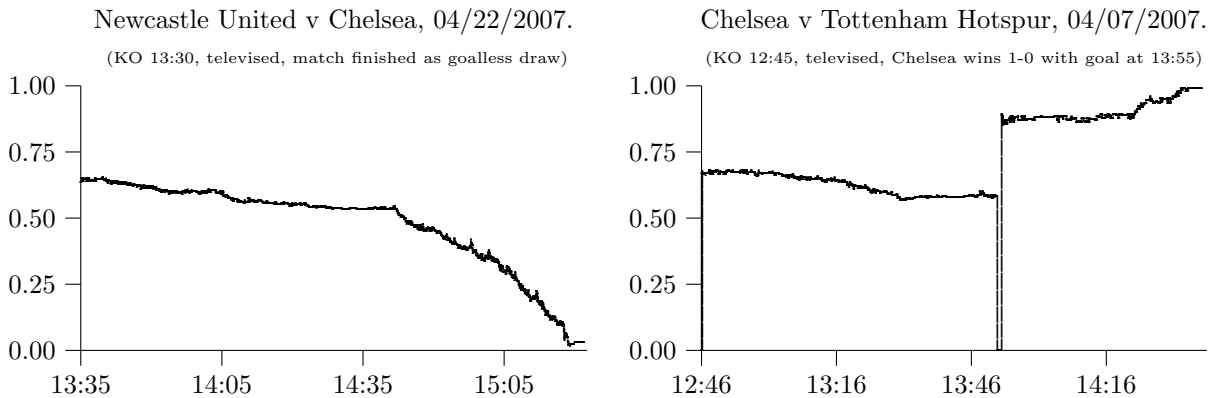

Figure 5: Illustrative time-related drift: evolution of Chelsea-win probability.

---

[28]McHale and Scarf (2006) document differences between domestic and international soccer games, notably that the gap in quality between competing teams tends to be smaller for domestic matches.

The left panel relates to an encounter between Newcastle and Chelsea. Depicted is the in-running probability (as implied by the best available 'back' price) that Chelsea—strong favorites at the start of the match—will win. The match ended in a goalless draw, and note that the win probability clearly drifts downwards as the clock ticks down and the chances recede of Chelsea breaking the deadlock. The right panel relates to another match involving Chelsea, in this case a game at home to Tottenham Hotspur a few weeks earlier. As the first half slips away without a goal, the probability of a Chelsea win begins again to drift downwards. At 13:55, just after the second half commences, Chelsea scores to move into the lead. As expected, the probability of a Chelsea win jumps up in response to the goal. It continues from there to drift upwards over the remainder of the game. This upward drift, at least partially, will reflect rational updating to the closing window of time and any other minor news. But it may also reflect some sluggishness in updating to the goal, and therein lies the identification challenge. In cases such as this, we might consider modelling rational price movement as a way to identify possible drift associated with inefficiency. Modelling even time-related drift is not trivial, however; it will depend on various factors, including the current scoreline (e.g., the magnitude of any lead) and the phase of play considered. In the next section, we introduce a simple and clean identification strategy that allows us to sidestep this complication; we exploit the (virtually) newsless window provided naturally by the half-time interval.

4.2.2 An Identification Which Exploits the Half Time Interval

Concretely, we propose to focus on games where goals are registered on the cusp of half-time and examine the way that prices in these markets behave during the interval. The half-time break, where time-related drift cannot be present, implies a natural opportunity to test cleanly for news-related drift at the level of individual contracts.

Our data set contains 160 goals that arrive within five minutes of the end of the first half—henceforth 'cusp goals'.
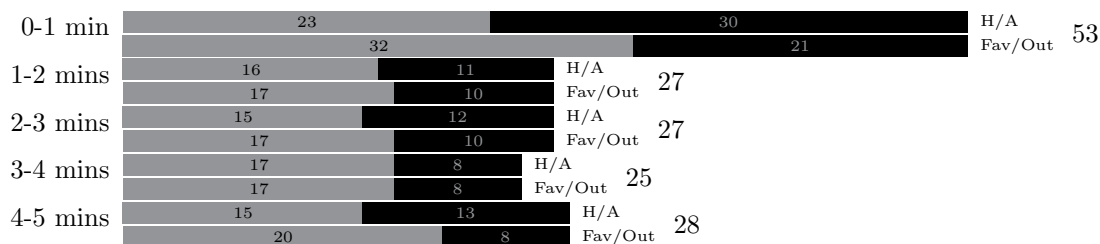
Figure 6: Cusp goals: 160 goals scored within five minutes of the start of half time.

Figure 6 takes a closer look at the distribution of such goals.[29] Home goals (H) account for 76 of the cusp goals; the other 84 therefore away goals (A). Favourites (Fav) score 103, and outsiders (Out) the remaining 57. Looking more closely at goal timings, a promising number of goals are scored extremely close to the end of first half; 53 occur in the final minute of play (within sixty seconds of the precise end of the first half of play), and a further 27 arrive in the penultimate minute.[30] This relative abundance of goals on the very edge of the break will be helpful for the purposes of our empirical strategy; the closer the goal to half time, the stronger the efficiency test.

---

[29]A small number of matches feature two cusp goals and we include in analysis only the later goal (that is, the

Implied probability of Manchester United win.

US$ volume traded each minute.
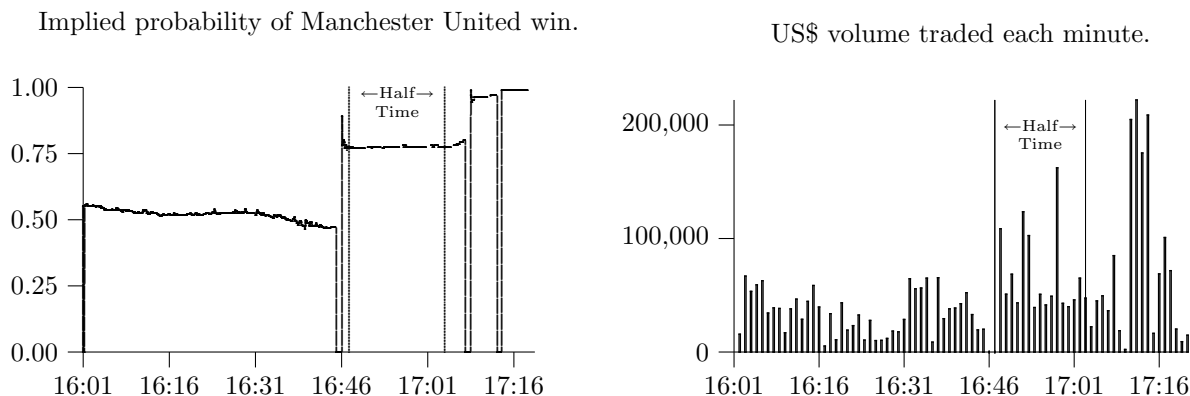
Figure 7: Visual evidence for efficient updating: Tottenham vs. Manchester United.

A visual look at the price series for a few of these matches is suggestive of efficient updating. Consider, first, Figure 7, in which Tottenham Hotspur plays at home to Manchester United. This match kicks off just after 16:00 and the plot in the left hand panel shows the probability of a Manchester win, as implied by the best Betfair back price. This probability is 56% at the beginning of the game (Manchester being favorites to win) but begins to drift downwards as the first half progresses without a goal. By the 44th minute it has fallen to under 50%. Then, right at the end of the first half, Manchester scores to take the lead and the market is suspended briefly. When it reopens moments later, the probability has jumped up to 77%. Almost immediately the whistle blows for half-time. Over the fifteen-minute interval that follows, the implied probability appears to remain remarkably constant at this 77% level, suggesting that updating to the goal was immediate and complete. A legitimate concern might be that such evidence for efficiency is an artefact of our half-time identification strategy: perhaps prices appear not to continue to update over half-time only because trading interest drops off during the break. In the right hand panel we report trading activity. Certainly for this game, the market is actively traded throughout half-time. In fact trading interest appears to step up somewhat during the interval in play.

For contrast, Figure 8 illustrates the case of an upset. Here, the ex-ante favorite concedes a goal just before half-time. The post-goal probability (again, as implied by the best Betfair back price) seems somewhat more volatile but still there appears no obvious trending. Meanwhile, trading during the half-time break is again heavy. Looking across all matches in our sample an average of $527 is traded per second of play and $319 is traded per second of half time. In games featuring a goal just before half time, we see an elevated half time trading volume of $552 per second.

The apparent lack of half-time trending in such diagrams, despite heavy trading during the break, constitutes prima facie evidence that the market is semi-strong form efficient. The rest of this section implements regression analysis to test this formally. Broadly, our strategy here is to test whether a goal occurring right at the end of the first half period creates any drift in prices observed over the half-time interval.

An important first step in implementing our test for semi-strong market efficiency is to construct an appropriate model of prices over the half time break. Suppose a security price at time t, P_t, can

---

goal scored closest to the end of the first half).

[30]We obtain precise timings for the end of the first half from Betfair, which records these for the purpose of closing its 'half-time score' betting markets.

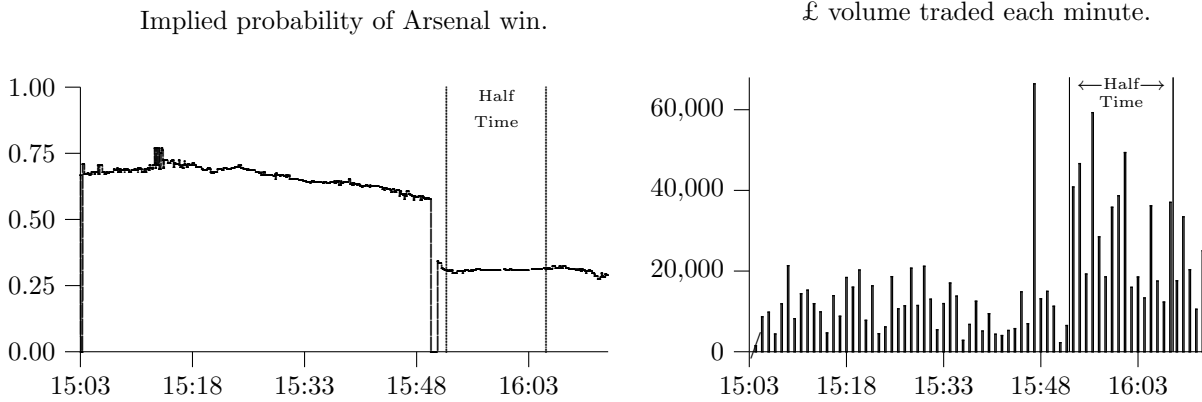Implied probability of Arsenal win.

£ volume traded each minute.

Figure 8: Further visual evidence for efficient updating: Arsenal vs. West Ham.

be written as the rational expectation of some fundamental value V^* conditional on information I_t available at t. Then we have $P_t = E_t[V^*]$, and $P_{t+1}, = E_{t+1}[V^*]$, which means, using the Law of Iterated Expectations, that $E_t[P_{t+1} - P_t] = E_t[E_{t+1}[V^*] - E_t[V^*]] = 0$. Thus, market efficiency means that realized changes in prices are unforecastable given information I_t. In other words, given currently known information, the expectation of an asset's price next period is simply its current price. Where this information set consists of historical prices then this is the martingale hypothesis (Le Roy 1989) and forms the basis of a test for weak-form efficiency. To implement this test would involve running regressions to ascertain whether or not current prices can be forecast on the basis of lagged prices. To test for semi-strong form efficiency (SSFE), which encompasses weak form efficiency but requires additionally that any public information is immediately incorporated in price, the information set comprises all relevant information that is publicly available at t. In our context, this means testing whether or not during the half time interval current prices can be forecast on the basis of lagged prices and checking to see whether this forecastability is impacted by the arrival of a goal on the cusp of half time.

Pooling the half-time price data across our full sample of matches, we could estimate for each contract type $m \in \{h, a, d\}$ the following second-order autoregression:

$$pr_{c,m,t} = \theta_0 + \theta_1 pr_{c,m,t-1} + \theta_2 pr_{c,m,t-2} + \varepsilon_{c,m,t}. \tag{1}$$

where pr denotes the volume-weighted average price and the subscripts $m \in \{1, \ldots, M\}$, $c \in \{1, 2, 3\}$, and $t \in \{1, \ldots, T\}$ denote the particular match, contract traded (Home Win, Away Win, Draw), and second of time considered. For each contract type, we have a panel data set comprised of prices for the contract type in question over our $M = 1206$ soccer matches. It makes sense to use panel estimation whenever there might be similarities or links between the processes generating the data in the different groups (here matches). In the presence of such links, combining the data can improve the efficiency of the estimation procedure. We do, however, run separate panel regressions for each contract type. Arbitrage opportunities make it very likely that the prices associated with the three match outcome contracts (for any given match) will be correlated (we can expect that, at any point during the match, the prices for the Home Win, Away Win and Draw contracts will imply probabilities that sum to about one) and running separate panel regressions for each contract type avoids potential complications due to this form of correlation in the data. Of course we can

17

also expect the prices for an individual contract to be correlated over time, but the inclusion of lagged dependent variables deals with this issue.

Under the null hypothesis of market efficiency, all coefficients in 1 would be zero except for the first lag of price, which would be significant and take a value of unity:

$$\mathbf{H_0}: \ \theta_0 = \theta_2 = 0, \theta_1 = 1. \tag{2}$$

Note that the model in 1 constrains the regression coefficients to be the same for all units of the panel; each soccer match conforms to exactly the same data generating process. Although the efficient markets hypothesis does constrain the parameters of the model to be constant across matches, it seems somewhat restrictive to impose parameter homogeneity at the initial estimation stage. Allowing for parameter heterogeneity in estimation seems prudent, because if the homogeneity assumption is wrong, the model will be misspecified and any testing adversely affected. In light of this concern, we prefer to work with a more generalized version of the model:

$$pr_{c,m,t} = \theta_{0,c,m} + \theta_{1,c,m}pr_{c,m,t-1} + \theta_{2,c,m}pr_{c,m,t-2} + \varepsilon_{c,m,t}. \tag{3}$$

Here the parameters of the model, θ_i, $i = 1, 2, 3$, are allowed to vary across matches (we are already allowing the parameters to vary between contracts by modelling the Home Win, Away Win and Draw markets separately). The model in (3) allows matches in which goals occur immediately before half time to differ from other matches, because each individual match will have its own set of parameters (e.g. $\theta_{0,h,1}, \theta_{1,h,1}, \theta_{2,h,1}$ for the Home Win contract in match 1). If our testing accepts H_0 over the entire panel (implying that the restrictions $\theta_0 = \theta_2 = 0, \theta_1 = 1$ are valid across the panel), then we are able to conclude in favour of market efficiency: half time prices reflect all relevant information, even when major news arrives on the cusp of half time. If, on the other hand, market efficiency is rejected, then we can investigate relevant groupings of matches, such as matches with cusp goals, to determine whether particular types of matches drive any rejection of the null hypothesis.

The next step is to transform (3), as is standard for unit root testing, into:

$$\Delta pr_{c,m,t} = \theta_{0,c,m} + (\theta_{1,c,m} - 1 + \theta_{2,c,m})pr_{c,m,t-1} - \theta_{2,c,m}\Delta pr_{c,m,t-1} + \varepsilon_{c,m,t} \tag{4}$$
$$= \gamma_{0,c,m} + \gamma_{1,c,m}pr_{c,m,t-1} + \gamma_{2,c,m}\Delta p_{c,m,t-1} + \varepsilon_{c,m,t}. \tag{5}$$

where $\Delta pr_{c,m,t} = pr_{c,m,t} - pr_{c,m,t-1}$, and $\gamma_{0,c,m} = \theta_{0,c,m}$, $\gamma_{1,c,m} = \theta_{1,c,m} - 1 + \theta_{2,c,m}$ and $\gamma_{2,c,m} = -\theta_{2,c,m}$. With respect to this transformed model, the null hypothesis of market efficiency implies:

$$\mathbf{H_0}: \ \gamma_{0,c,m} = \gamma_{1,c,m} = \gamma_{2,c,m} = 0, \tag{6}$$

which is a standard F-test of joint significance of regression coefficients. We note that the null hypothesis for the standard Augmented Dickey-Fuller (ADF) unit root test would be simply $\gamma_{1,c,m} = 0$, and hence because the null hypothesis implies non-stationarity, we must compare the F-test of the semi-strong form efficiency hypothesis to a non-standard distribution. Patterson (2000) contains critical values for the joint test involving $\gamma_{0,c,m}$ and $\gamma_{1,c,m}$ for a particular contract and match, but not the joint test including the lagged difference term, nor a panel variant of the test. Our data is also high frequency and as is often the case in such data series, our residuals are very leptokurtic compared to the standard normal distribution assumed for generating p-values on

	Maddala, Wu			Im, Peseran, Shin		
	home	away	draw	home	away	draw
1% Critical Value $\left(\text{standard } \chi^2_{2M}\right)$	16413 (2576.5)	16537. (2576.5)	16592. (2576.5)	6.13	6.13	6.13
5% Critical Value $\left(\text{standard } \chi^2_{2M}\right)$	4752.6 (2527.4)	4788.6 (2527.4)	4804.6 (2527.4)	4.51	4.51	4.51
Test statistic	8004.7	7853.6	8355.4	6.44	6.92	6.78

Table 2: Results from statistical efficiency test applied to first five minutes of half time, with the standard χ^2_{2M} critical values for the M-W test reported beneath the simulated critical values.

which to calculate one of our two tests.[31] Therefore, we must simulate critical values. We provide details of our simulation exercise in the Appendix. We simulate critical values for two common variants of panel unit root tests: the Maddala and Wu (1999) (M-W) and Im et al. (2003) (IPS) tests. The IPS test aggregates based on the ADF test statistics for each individual time series, whereas the M-W test is a Fisher test aggregating the p-values from the individual tests. The IPS test statistic is:

$$IPS_c = \frac{1}{M} \sum_{m=1}^{M} Z_{m,c},\tag{7}$$

where $Z_{m,c}$ is the ADF test statistic for match m and contract c. Critical values for this test are simulated by Im, Pesaran and Shin, but not for the specific case we consider here, hence the need to simulate critical values. The Maddala-Wu test statistic is:

$$MW_c = -2 \sum_{m=1}^{M} \ln p_m \sim \chi^2_{2M},\tag{8}$$

where p_i is the p-value for the ADF variant test, or SSFE test, in the ith time series of the panel. The test statistic is χ^2-distributed with $2M$ degrees of freedom for the standard unit root case when errors are assumed normally distributed. We simulate for the case where errors are t-distributed to reflect the leptokurtic nature of errors in high-frequency datasets, and compare the generated critical values with the standard critical values (provided in parentheses in Tables 2 and 3).

Results for the test applied to the first five minutes of half-time prices are shown in Table 2 and results for the first ten minutes of half time appear in Table 3. In both cases we are able to compare (for the M-W test) the critical values both from using the standard assumption of normality (implying a χ^2_{2M} distribution for the test statistic), and from our simulations assuming our errors are t-distributed with three degrees of freedom (see again the Appendix for more details on our simulations). Simulating our market efficiency test for individual time series reveals that critical values must be somewhat more generous in order to ensure appropriate test properties (as is the case with simple unit root tests), particularly at the 1% level, reflecting the leptokurtic nature of the t-distribution relative to the normal distribution.

Overall, we reject semi-strong form efficiency at the 1% level, but not at the 5% level. We conclude that the statistical efficiency of the markets at half time is only somewhat supported. If

---

[31]We conducted Doornik and Hansen (2008) normality tests on our prices for each match and found that (results tabulated in the Appendix) in 97–100% of matches the test fails, suggesting that indeed the normality assumption is not an appropriate one.

	Maddala, Wu			Im, Peseran, Shin		
	home	away	draw	home	away	draw
1% Critical Value $\left(\text{standard } \chi^2_{2M}\right)$	16592 (2576.5)	16371 (2576.5)	16495 (2576.5)	16.77	16.77	16.77
5% Critical Value $\left(\text{standard } \chi^2_{2M}\right)$	4805.9 (2527.4)	4741.8 (2527.4)	4777.8 (2527.4)	3.73	3.73	3.73
Test statistic	6988.1	7131.3	7505.11	9.59	10.68	9.78

Table 3: Results from statistical efficiency test applied to first ten minutes of half time, with the standard χ^2_{2M} critical values for the M-W test reported beneath the simulated critical values.

we had used the standard critical values, the rejection of the null hypothesis would have been much more emphatic for both the five- and ten-minute regressions, but as argued earlier, a leptokurtic distributional assumption is much more appropriate for high-frequency financial data series.

Most meaningful ultimately is not whether markets are efficient in a statistical sense, but whether customers could trade profitably on any inefficiency. Hence, we focus in the remainder of this section on investigating economic efficiency, testing whether customers could make positive returns over the half-time interval by exploiting any systematic drift in prices (related to a cusp goal or otherwise). We evaluate two hypothetical trading strategies designed to exploit potential half-time price drift:

- Trading strategy A: backing (buying) a particular match outcome at the start of the half time interval and laying this (selling it back to the market) 5 (or 10) minutes later. This would exploit any systematic downward drift in odds during the break, perhaps reflecting an initial underreaction to the arrival of a goal just before the half time interval.

- Trading strategy B: laying (selling) a contract at the start of the half time interval and backing this 5 (or 10) minutes later. This strategy would exploit any systematic upward movement in odds during the interval, perhaps reflecting an initial overreaction to a goal at the end of the first half.

The potential profitability of each strategy can be investigated using a difference in means test. Denote as $p_{m,b,i}$ the best back price for a particular outcome in minute m of the half time interval in match i, the best lay price as $p_{m,l,i}$, and their respective means across all matches in the sample as $\bar{p}_{m,s}$, where $s \in (b, l)$ is the 'side' of the market (back or lay). Note that at any point in time, the best available price to back a particular outcome at that moment in time (e.g., decimal odds of 9) will be below the best available price to lay the same outcome (e.g., decimal odds of 9.2). This must be the case in any limit order book, otherwise the exchange could immediately match some of the orders in the book by crossing trades at prices between 9 and 9.2. Consequently, it will only be possible to profit from strategy A if Then, to test the profitability of, for example, backing an outcome in the first minute of half time and laying it in the fifth minute (Strategy A), a suitable difference in means test would calculate the t-statistic:

$$t = \frac{\bar{p}_{1,b} - \bar{p}_{5,\ell}}{\sigma_{\bar{p}_{1,b} - \bar{p}_{l,\ell}}}. \tag{9}$$

The test is a paired t-test (comparing the prices at minute 1 with those at minute 5 for each match), and so $\sigma_{\bar{p}_{m,b} - \bar{p}_{m,\ell}}$ is simply the standard deviation of the difference between the relevant back and lay prices divided by the square root of the number of matches considered. If trading

	Home Win					Away Win					Draw				
	Average price			Test		Average price			Test		Average price			Test	
	Back	Lay				Back	Lay				Back	Lay				
Interval minute	1	5	10	5	10	1	5	10	5	10	1	5	10	5	10
All	8.47	10.1	11.1	-3.40	-2.85	17.3	23.6	22.9	-4.40	-4.90	6.47	6.92	6.99	-5.99	-5.79
Vol> 5	8.15	9.76	10.8	-3.22	-2.69	14.9	20.1	19.5	-3.63	-3.96	6.32	6.63	6.79	-6.86	-6.12
Vol> 10	7.34	8.97	9.77	-2.91	-2.43	12.8	16.6	17.5	-3.37	-3.75	6.23	6.58	6.65	-6.60	-5.86
Vol> 20	5.93	6.90	7.42	-2.92	-2.39	11.2	14.6	15.5	-2.93	-3.34	6.16	6.51	6.59	-6.01	-5.43
Vol> 50	4.90	5.28	5.56	-4.79	-3.70	8.64	10.2	11.9	-2.69	-2.32	5.51	5.78	5.85	-4.27	-3.89

Table 4: Difference in means test for the profitability of trading strategy A: backing an outcome at the start of the half time interval and laying it (selling) 5 or 10 minutes later.

	Home					Away					Draw				
	Average price			Test		Average price			Test		Average price			Test	
	Lay	Back				Lay	Back				Lay	Back				
Interval minute	1	5	10	5	10	1	5	10	5	10	1	5	10	5	10
All	11.1	8.55	8.77	-2.89	-2.93	23.0	18.1	18.4	-4.10	-3.79	7.14	6.44	6.51	-4.58	-4.08
Vol> 5	13.2	9.89	10.1	-2.50	-2.63	21.7	19.0	19.1	-3.54	-3.17	8.45	7.62	7.60	-3.95	-3.69
Vol> 10	13.5	9.84	10.0	-2.42	-2.59	20.5	18.0	18.2	-3.10	-2.61	8.38	7.67	7.67	-4.21	-4.54
Vol> 20	12.6	9.22	9.41	-2.09	-2.30	20.0	17.3	17.4	-2.89	-2.56	7.57	6.99	7.01	-6.10	-5.65
Vol> 50	11.9	8.15	8.38	-1.90	-2.12	18.4	15.8	15.9	-2.29	-1.92	5.98	5.68	5.53	-4.63	-4.29

Table 5: Difference in means test for the profitability of trading strategy B: laying an outcome at the start of the half time interval and backing it 5 or 10 minutes later.

strategy A is profitable, we would expect $p_{1,b} - p_{5,\ell}$ (or $p_{1,b} - p_{10,\ell}$) to be significantly positive: the return from backing the outcome in question in the first minute of half time must be greater than the exposure required to cover this position (sell it back into the market) in minute five (or ten) of the break. Since our only concern is whether $p_{1,b} - p_{5,\ell} > 0$ (or $p_{1,b} - p_{10,\ell} > 0$) the test is one tailed.

We report results from this test in Table 4. The first row gives results for all matches for which data was available (1062 matches for the 5 minute test, 1055 for 10 minutes). In the rows below this, we restrict our attention to gradually smaller subsections of the available matches, dictated by the volume available to trade at the start of half time. Disaggregating by available volume will provide a finer picture of the viability of our hypothetical trading strategies for different sizes of bet. In the second row, we omit from consideration all matches where as half time began less than $ 5 was available to back; this leaves 653 matches. The third row reports results when we omit those matches where less than $ 10 was available to bet at the best available back price; this leaves 553 matches. Finally, in the fourth and fifth rows we omit matches where less than $ 20 and less than $ 50 was available to bet at the best available back price; this leaves 478 and 372 games, respectively.

Although the t-statistics in Table 4 are quite large, they are negative, implying a clear rejection of the null hypothesis that the difference in means is non-positive. This result implies that Strategy A is not profitable: backing the event at the start of half time and laying after 5 or 10 minutes does not yield a positive expected return.

Trading strategy B involves laying an event at the start of half-time and subsequently backing the event (after 5 or 10 minutes). In this case, the trader makes a positive return only if $p_{5,b} - p_{1,\ell} > 0$ or $p_{10,b} - p_{1,\ell} > 0$. Hence again the test is one-tailed and the critical value is 1.96. The results are reported in Table 5.

As before, the first row of this table considers all matches for which data as available, whilst the rows below this consider progressively smaller subsets of matches according to the availability of volume at the best available lay price. In the second row, matches are omitted where as half time began less than $ 5 was available to lay (leaving 644 matches). In the third, fourth, and fifth

rows, matches are excluded where less than $ 10, $20, $50 was available to lay (leaving 544, 471, and 367 matches, respectively).

The results in Table 5 confirm clearly that trading strategy B is also unprofitable: the average decimal odds that must be offered for the lay trade is always strictly greater than those available for the back trade, implying that on average this strategy yields a negative return.

Overall then, our half-time identification yields conclusive evidence that Betfair markets are economically efficient: prices impound news so rapidly and completely that it is not possible to exploit any systematic drift over the half time interval for profitable trading. This is so even where major news (a goal) arrives just seconds before the end of first half play.

4.2.3 Testing More Directly for Efficiency In-Play: Two Complementary Approaches

An attractive feature of the half-time strategy of the previous section is its cleanness. A potential concern is that our results concerning efficiency could be specific to the break in play. There are many conceivable reasons why this might be so. Perhaps different types of trader are active during the half-time interval; perhaps major news interacts with more minor news during minutes of play, which might complicate updating. In this section, we develop two complementary strategies for testing the market's ability to update to the news of a goal whilst the match is in progress.

(I) Testing for Post-Goal Drift at the Aggregate Level.

Although individual series will necessarily display drift, we should fail to find evidence of drift at the aggregate level if the market is efficient. The calibration tests now common to the prediction markets literature can be seen as special cases of a test for drift in average prices; a standard calibration test asks whether the current price is the best forecast of the final price. In common with standard calibration tests, and as explained in our introduction, a test for drift in average in-play prices can only deliver meaningful results regarding efficiency when applied to a sufficiently large sample.

With 1,206 matches, we can hope to perform a meaningful test along these lines. For ease of benchmarking with previous work (Gil and Levitt 2007) we do this deploying a regression of the form:

$$p_{m,c,w,t} = \beta_0 + \sum_{g=-14, g \neq 0}^{15} \beta_g Goal_{m,c,w,t+g} + \varepsilon_{m,c,w,t}, \quad \varepsilon_{m,c,w,t} \sim IID\left(0, \sigma^2\right), \qquad (10)$$

The model is estimated on a panel dataset consisting for each match and market of the weighted-average prices observed over the thirty minutes surrounding a goal (fifteen minutes either side of this news). p denotes the volume-weighted average price and the subscripts m, c, w, and t denote the particular match, contract traded (we consider Home Win and Away Win contracts), goal 'window', and minute of time considered.[32] $Goal$ is an indicator variable corresponding to a particular minute which is equal to unity when a goal is scored in favour of the contract, and zero otherwise. The matches in our dataset yield a total of 3,078 goals, and we are able to consider 2,528 of these in our panel dataset. (A large number occur either very near the beginning or towards the end of a match, leaving insufficient observations in the pre- or post-goal period to support

---

[32]Gil and Levitt (2007) also demean their price data; since this is a linear transformation its effect is simply captured here in the constant coefficient, β_0. Gil and Levitt omit this term.

estimation.) We estimate (10) using pooled OLS. The semi-strong form efficiency hypothesis takes the form:

$$\mathbf{H}_0: \ \beta_1 = \beta_k, k > 1. \tag{11}$$

We report our results in the left panel of Table 6) and plot the coefficients on goals in Figure 9 along with 95% confidence bands.[33]
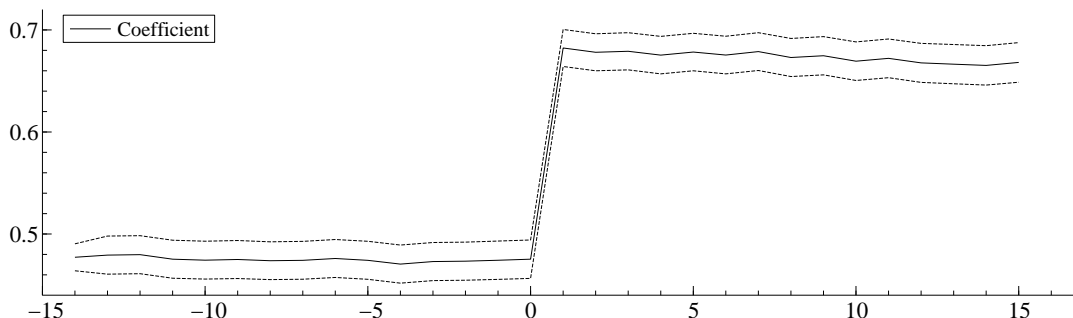

Figure 9: Testing for semi-strong form efficiency during minutes of play, by looking for inefficient post-goal drift in *average* in-play prices. This figure presents the β coefficients from model 10, where the time of a goal is normalized to $t = 0$.

In common with Gil and Levitt (2007) we find an initial jump of around 0.2 in magnitude. However, in contrast to their study, we are unable to find evidence of post-news drift. Coefficients for post-goal minutes two to fifteen do not fall outside the 95% confidence bands associated with the coefficient on the first post-goal minute. Hence, it seems that the coefficients are not drifting. Supporting this conclusion further, F-tests of the null hypothesis that $\beta_1 = \beta_k$, $k > 1$ are never rejected at the 5% level. Moreover, and as discussed already in the previous section, it is important to appreciate the implications of our large sample size for the appropriate choice of critical test values. When we consider individual minutes and seconds from our sample of 1,206 matches, we again generate vast datasets; the sample we employ in this testing section comprises 64,450 observations. For a sample size of 64,450, the appropriate p-value of 0.000142.[34] This translates into a critical t-statistic of 3.8, as opposed to the familiar value of 2.

The post-goal markets in our sample might be viewed as incorporating a 'structural break' in the parlance of econometricians; a clear discontinuity is observed in relation to how agents form beliefs in the market, and hence the realisation of the market price. Econometric methods for uncovering structural breaks abound in the literature; the method proposed by Bai and Perron (2003) is perhaps the best known. Inserting a dummy variable for (almost) every observation and testing their statistical significance would be consistent with an alternative strategy of 'dummy saturation' advanced by Hendry et al. (2008). Based on these ideas, the efficiency test just presented might be recast in terms of a test for a structural break at the time of the goal, followed by stability in average price. The significant post-goal dummy variables strongly suggest that a

---

[33]Again, our sample size suggests that confidence bands well above 95% would be most appropriate but we report at the 95% for more explicit calibration to Gil and Levitt (2007).

[34]The reader is referred to the arguments of the preceding section for clarification of the appropriate rule for computing sample-adjusted critical values.

Variable	In-Play Efficiency Test based on model (10)					In-Play Efficiency Test based on model (12) .		
	Coef.	Std. Err.	t	95% Conf. Int.		Coef.	Std. Err.	t
shift			22.17	0.1870	0.2233	0.2051	0.0093	22.17
$goal_1$	0.2051	0.0093	21.68	0.1828	0.2191			
$goal_2$	0.2009	0.0093	21.74	0.1837	0.2201	-0.0042	0.0090	-0.47
$goal_3$	0.2019	0.0093	21.16	0.1798	0.2165	-0.0032	0.0090	-0.36
$goal_4$	0.1981	0.0094	21.48	0.1828	0.2195	-0.0070	0.0091	-0.77
$goal_5$	0.2011	0.0094	21.03	0.1797	0.2167	-0.0040	0.0091	-0.44
$goal_6$	0.1982	0.0094	21.32	0.1831	0.2201	-0.0069	0.0091	-0.76
$goal_7$	0.2016	0.0095	20.55	0.1771	0.2145	-0.0035	0.0092	-0.39
$goal_8$	0.1958	0.0095	20.67	0.1788	0.2162	-0.0093	0.0092	-1.01
$goal_9$	0.1975	0.0096	19.91	0.1733	0.2111	-0.0076	0.0093	-0.82
$goal_{10}$	0.1922	0.0097	20.12	0.1760	0.2140	-0.0129	0.0094	-1.38
$goal_{11}$	0.1950	0.0097	19.54	0.1714	0.2096	-0.0102	0.0094	-1.08
$goal_{12}$	0.1905	0.0097	19.30	0.1700	0.2085	-0.0146	0.0095	-1.54
$goal_{13}$	0.1893	0.0098	19.09	0.1687	0.2074	-0.0159	0.0095	-1.67
$goal_{14}$	0.1880	0.0098	19.30	0.1716	0.2104	-0.0171	0.0096	-1.79
$goal_{15}$	0.1910	0.0099	19.29	0.1710	0.2108	-0.0141	0.0096	-1.47
Constant	0.4772	0.0068	70.69	0.4640	0.4904	0.4772	0.0068	70.69

Table 6: Testing for semi-strong form efficiency during minutes of play, by looking for inefficient post-goal drift in in-play prices. The left panel of the table reports the β coefficients from model (10). The right hand panel reports the β coefficients from estimation of model (12).

structural break has occurred. Hence, a suitable respecification of model (10) would include a shift dummy that takes the value unity when $g > 0$, and omit $Goal_{m,c,w,t+1}$, resulting in:

$$p_{m,c,w,t} = \beta_0 + \beta_s Shift_{m,c,w,t} + \sum_{g=-14, g\neq 0, g\neq 1}^{15} \beta_g Goal_{m,c,w,t+g} + \varepsilon_{m,c,w,t}, \quad \varepsilon_{m,c,w,t} \sim IID\left(0, \sigma^2\right),$$
(12)

where $Shift_t = 1_{\{t \geq 1\}}$. A test based on equation (12) is entirely equivalent econometrically to our test based on equation (12). However, we now need deploy only a simple t-test for variable significance (or F-test of joint significance) to investigate efficiency. Because the structural break is captured by the β_s coefficient, any drift in average prices subsequent to a goal will manifest in significant β_g coefficients. Hence this test is more transparent and also splits the initial post-news price jump from any subsequent drift.

The final panel of Table 6 reports the results from testing based on equation (12), and it can be seen from the Shift coefficient that $\beta_s = \beta_1$, confirming that the two approaches give equivalent indications of the average size of the initial post-goal jump. Furthermore, we note also that the highest t-statistic is -1.79, which is insignificant at the 5% level and highly insignificant at our sample-adjusted significance level (p-value 0.000142). As a result, we are again unable to reject the hypothesis that Betfair markets are semi-strong form efficient during minutes of play.

Thus our results from this section would appear to support the view that these markets also update efficiently to major news during minutes of play. We are unable to find significant evidence of post-goal drift, in contrast to Gil and Levitt (2007), and despite deploying tests comparable to the test in their paper. As Gil and Levitt have a relatively small sample (fifty matches) drawn from considerably thinner betting markets, we suggest that there are two possible explanations for their apparent finding of some inefficiency (post-goal drift). One possibility certainly is that the markets they study are inefficient, perhaps because of cognitive biases on the part of traders or simply because these markets are very thin (low volumes, infrequent trades). An alternative explanation might be that the implemented test for efficiency, being a test for 'average' in-play drift in binary assets, cannot yield meaningful results when applied to small samples.[35]

(II) Modelling 'Efficient' In-Play Drift in Individual Matches

In this section, we develop a simple model for generating minute-by-minute probabilities of final match outcomes. We use this model to construct an 'efficient' purely time-related drift for the 547 English Premiership matches in our sample. Then, for each match, we compare our synthetic drift to the drift actually observed in Betfair implied probabilities in order to draw inferences regarding the extent to which observed Betfair price drift in individual games reflects the market's efficient adjustment to the ticking down of playing time.

In order to construct efficient time-related drift, we begin by looking for a well-fitting model of goal arrival and using this to back out minute-by-minute match outcome probabilities. Our

---

[35]We should record a few reservations about the tests carried out in this section from an econometric point of view. If these market are efficient, then prices will follow a martingale, and so the lagged price, if included in our regressions, would have have a coefficient of unity and be strongly significant. Even without efficiency, one suspects that lagged dependent variables will be significant. By omitting this coefficient, test equations could be misspecified, and standard errors as a result may be unreliable. Heteroskedasticity and autocorrelation (HAC) robust standard errors can be employed to attempt to correct for this, and are applied to our own test equations. However, procedures such as using robust standard errors attack only the symptoms of misspecification, and not the cause. Inference may suffer as a result.

approach is to combine two statistical models in an intuitive way. Firstly, we utilize the Bivariate Poisson model developed by Karlis and Ntzoufras (2003). We must break the match-level analysis of Karlis and Ntzoufras (2003) into smaller intervals in order to uncover the in-match variation, and in particular drift, that we are interested in. It is well documented that at the match level, goal arrival is Poisson distributed, and examination of our Premiership match outcome database reveals that the Poisson assumption is also valid for minute-by-minute action: across all English Premiership matches from the 2006–07 and 2007–08 seasons, goal arrival in a given minute follows a Poisson process with a mean of 0.016 goals per team per minute and a variance also equal to 0.016. Thus, extending the Karlis-Ntzoufras framework for modelling minute-long intervals for each match is valid statistically.[36] This provides estimates of the probability of a Home Win, Away Win, and Draw for a given minute interval. However, we require probabilities of eventual match outcomes, as opposed to the outcomes for each minute. As such, we take these estimated probabilities, and treat these as the $p = 3$ events repeated n times in a multinomially distributed event. The n events are the remaining minute intervals before the end of the match, and thus our approach provides, for every minute in the match, a final match-outcome probability for each market. As a number of plots will reveal, the resulting model delivers a rather accurate prediction of the match outcome probabilities implied by Betfair prices.

Formally, each team's rate of goal arrival in each minute (m) of a match (i) is a Poisson-distributed variable with intensity parameter $\lambda_{i,m}$. We specify three random variables, $X_{1,i,m}$, $X_{2,i,m}$ and $X_{3,i,m}$ as independently Poisson distributed with parameter $\lambda_{\kappa,i,m}$, $\kappa = 1, 2, 3$. Then, defining $X_{i,m} = X_{1,i,m} + X_{3,i,m}$ and $Y_{i,m} = X_{2,i,m} + X_{3,i,m}$, $X_{i,m}$ and $Y_{i,m}$ jointly follow a bivariate Poisson distribution $BP(\lambda_{1,i,m}, \lambda_{2,i,m}, \lambda_{3,i,m})$ with density function:

$$
\begin{aligned}
\mathsf{P}_{X_{i,m},Y_{i,m}}(x_{i,m}, y_{i,m}) &= \mathsf{P}\left(X_{i,m} = x_{i,m}, Y_{i,m} = y_{i,m}\right) \\
&= \exp\left\{-\left(\lambda_{1,i,m} + \lambda_{2,i,m} + \lambda_{3,i,m}\right)\right\} \\
&\quad \times \frac{\lambda_1^{x_{i,m}}}{x_{i,m}!} \frac{\lambda_2^{y_{i,m}}}{y_{i,m}!} \sum_{k=0}^{\min(x_{i,m},y_{i,m})} \binom{x_{i,m}}{k} \binom{y_{i,m}}{k} k! \left(\frac{\lambda_{3,i,m}}{\lambda_{1,i,m}\lambda_{2,i,m}}\right)^k.
\end{aligned}
\tag{13}
$$

Then $\mathsf{E}(X_{i,m}) = \lambda_{1,i,m} + \lambda_{3,i,m}$, and $\mathsf{E}(Y_{i,m}) = \lambda_{2,i,m} + \lambda_{3,i,m}$, while $\mathsf{Cov}(X_{i,m}, Y_{i,m}) = \lambda_{3,i,m}$. Hence $\lambda_{3,i,m}$ captures the dependence between the goal arrival rates of the two teams competing in a match. With $\lambda_{3,i,m} = 0$, then (13) reduces to two independent Poisson distributions—a double Poisson distribution. Following Karlis and Ntzoufras (2003), we suggest as a natural interpretation of the parameters that $\lambda_{1,i,m}$ and $\lambda_{2,i,m}$ reflect the 'net' scoring ability of each team whereas $\lambda_{3,i,m}$ reflects match conditions (e.g. the speed of the game, the weather, the pitch quality, or other conditions at the ground). To capture additional factors that might affect the goal arrival rates of the teams, vectors of explanatory variables $\mathbf{w}_{\kappa,i,m}$ specific to each match and minute can be introduced, and a bivariate Poisson regression can be run:

$$
\begin{aligned}
(X_{i,m}, Y_{i,m}) &\sim BP(\lambda_{1,i,m}, \lambda_{2,i,m}, \lambda_{3,i,m}), \\
\log(\lambda_{\kappa,i,m}) &= \beta_\kappa \mathbf{w}_{\kappa,i,m}, \qquad \kappa = 1, 2, 3.
\end{aligned}
\tag{14}
$$

---

[36]For modelling purposes, we assume that the market uses sample averages in forming its estimates of the 'injury time' likely to be added on by the referee at the end of each half of play.

The $\mathbf{w}_{\kappa,i,m}$ are specific to the match, minute and parameter being estimated, while β_{κ} is the vector of regression coefficients. The regression models for the three parameters are as follows:

$$\log(\lambda_{1,i,m}) = \mu_1 + \beta_{1,1}att_{h,i,m} + \beta_{2,1}def_{a,i,m}, \tag{15}$$

$$\log(\lambda_{2,i,m}) = \mu_2 + \beta_{1,2}att_{a,i,m} + \beta_{2,2}def_{h,i,m}, \tag{16}$$

$$\log(\lambda_{3,i,m}) = \mu_3 + \beta_3\mathbf{w}_{3,i,m}, \tag{17}$$

where $h_{i,m}$ denotes the home team in match i, minute m, and $a_{i,m}$ the away team. Scoring intensity, represented for the home and away teams by $\lambda_{1,i,m}$ and $\lambda_{2,i,m}$, is measured by the number of goals scored by each team in a given minute-long interval. As with a standard regression model, we estimate (15) to generate fitted values for $\lambda_{1,i,m}$. Because home goals in a game is used to estimate this, and as the two teams represented by dummy variables, the regression gives parameter estimates for home attacking ability ($\beta_{1,1}$) and away defensive ability ($\beta_{2,1}$). The $\lambda_{2,i,m}$ regression model is estimated based on the away goals scored in a game, giving equivalent parameter estimates for home defence and away attack. The covariance parameter $\lambda_{3,i,m}$ is allowed to vary with a number of factors captured in the vector of parameters $\mathbf{w}_{3,i,m}$. This vector consists of dummies for the teams competing in the match, the total goals scored by each team by minute m of the match, and the number of minutes elapsed, m.

By substituting into (13) the fitted values of $\lambda_{\kappa,i,m}$ estimated from (15)–(17), we can generate the probability at any minute in a match of a Home Win, Away Win or Draw for the next minute: $P(\text{Home Win}) = P(X_{i,m} > Y_{i,m})$, $P(\text{Draw}) = P(X_{i,m} = Y_{i,m})$, $P(\text{Away Win}) = P(X_{i,m} < Y_{i,m})$. Thus we have three outcomes (Home Win, Draw, Away Win), and a number of trials (intervals left until the end of the match). Assuming that the probabilities based on the model estimated on the given interval remain constant over the remaining intervals, one can then calculate probabilities for final match outcomes using the probability density for the multinomial distribution. If we have n trials with k possible outcomes, denoted x_k, where each outcome occurs with probability p_k such that $\sum_{i=1}^{K} p_i$, then:

$$f(x_1, \ldots, x_k; n, p_1, \ldots, p_k) = \begin{cases} \frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} p_i^{x_i} & \text{when } \sum_{i=1}^{k} x_i = n, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

This describes our simple strategy for generating 'efficient' minute-by-minute probability series for entire matches. We go on to apply this methodology to create predicted series for all 547 English Premiership matches in our sample.

Figure 10 shows the calibration of our model's predictions to observed match outcomes. The model provides a good fit overall. Meanwhile Figure 11 illustrates the fit between the in-running probabilities implied by Betfair prices and those predicted by our model for five matches selected at random from our sample. Some discrepancies between the two series are visible but overall the drift observed in these illustrative matches appears well approximated by our generated series; Figure 12 plots the difference between the two series in these four matches, and shows that any remaining drift in the Betfair series is negligible.

As a case study of our basic attempts to compare 'efficient' drift with Betfair price drift, we take a closer look at a match between Manchester United and Blackburn Rovers, which took place during the 2006–07 season. In Figure 13 we plot the probability of a Manchester United win, as implied by the observed Betfair price series, along with the predicted probability series from our
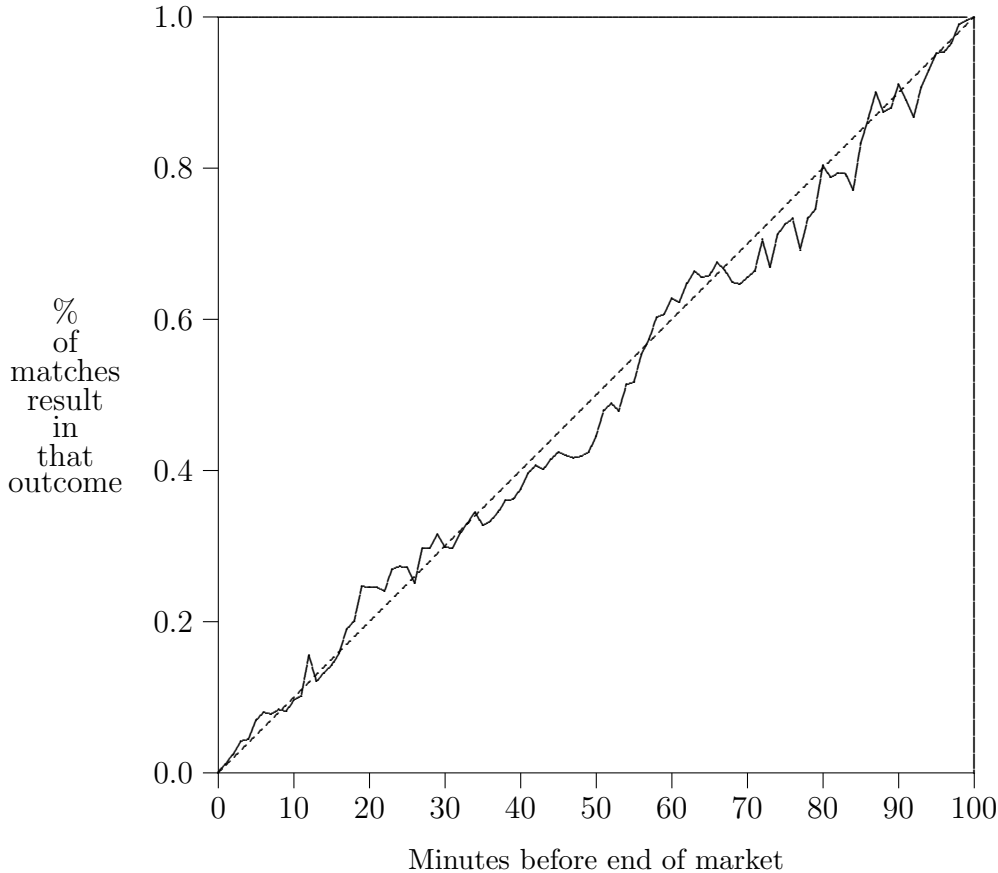
Figure 10: Calibration of the Poisson model to historical match data

model. As with a number of other matches, in the first half of the game, our predicted time series for the probability is somewhat above actual Betfair probabilities. Yet, as the right panel shows, during all phases in play our series is capturing the slope of the drift in Betfair prices almost perfectly. The first goal is scored by Blackburn at 1,700 seconds into playing time. In response to this goal, both series jump down (reflecting a drop in the probability of a Manchester win). Our estimated probability series drops less than the Betfair series, and the predicted downward drift remains marginally above the drift in Betfair probabilities as further goalless minutes pass, but again shares a similar slope. During the newsless half-time interval, unsurprisingly, we observe a relatively flat series both in observed and predicted probabilities. Following the break, at around 5000 seconds into the match, Manchester United levels the score and, following this, our predicted series and the actual Betfair series converge rapidly. By the time Manchester United scores again to take the lead (at around 5700 seconds into the game) our model's predicted 'efficient' trend and the observed trend in Betfair probabilities coincide.

A basic metric for closeness of fit between our Poisson model's predicted in-running time series and the observed Betfair series is a sum of squared deviations measure. We compute second-by-second deviations of Betfair probabilities from our model's predicted probabilities and aggregate to obtain a total per game sum of squared deviations:
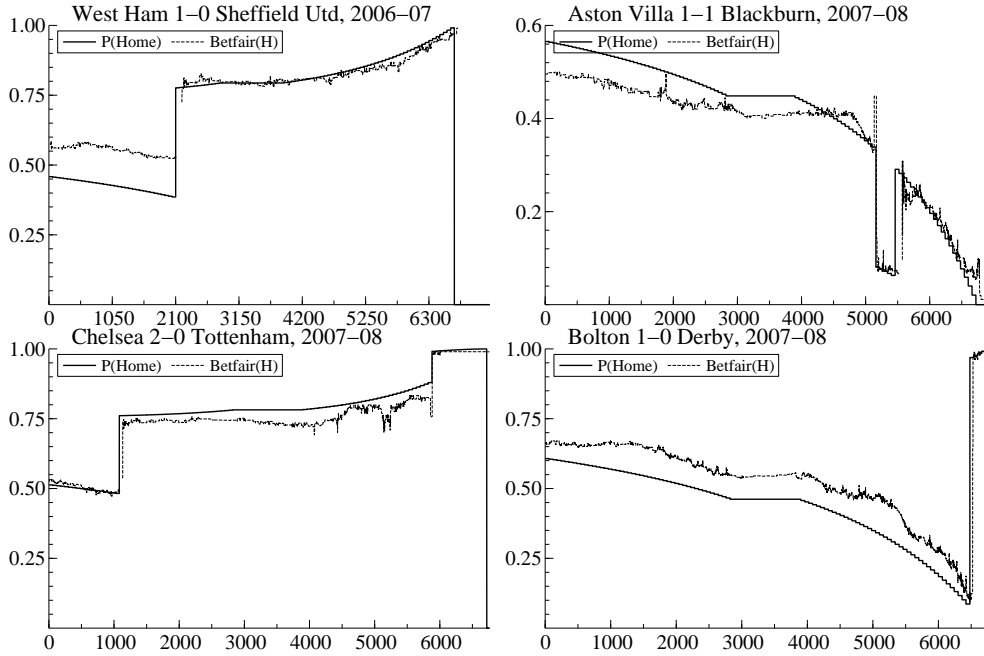
Figure 11: Plots of the Poisson-Multinomial minute-by-minute probability series and the Betfair implied probability series for illustrative English Premiership matches. The smoother, thicker plot is the Poisson series.

$$\mathsf{SSD} = \sum_{t=1}^{6540} \left(\widehat{P}_{Betfair,t} - \widehat{P}_{Poisson,t} \right)^2. \tag{19}$$

The number 6,450 reflects the average number of seconds in a match (computed by adding to the ninety minutes of normal playing time, an extra minute for first half injury time, a further three minutes for second half injury time, and finally, fifteen minutes for half time).[37] We report the summary statistics for this measure for our full sample of 547 English Premiership games in Table 7. Since this is a relative measure, and perhaps difficult to interpret in the abstract, we present in Table 8 the individual SSD values for the representative matches depicted in our plots in Figures 11 and 13.[38]

In summary, the simple modelling exercise carried out in this section gives preliminary support to the view that the majority of drift observed in Betfair markets during minutes of play might reasonably be explained by efficient updating to the passage of playing time.

---

[37]Hence, dividing each SSD by 109 would give an average (cumulated) squared deviation per minute.

[38]Some of these SSD values may appear rather high when one computes the per second average deviation, but follow-up investigation suggests that once outliers are removed (due to slight mistracking of goal times and some spurious events towards the end of games where market liquidity often dries up leaving unrepresentative prices in the market) the average difference between the series is very modest, as the plots, which we have selected for their representativeness, indicate.
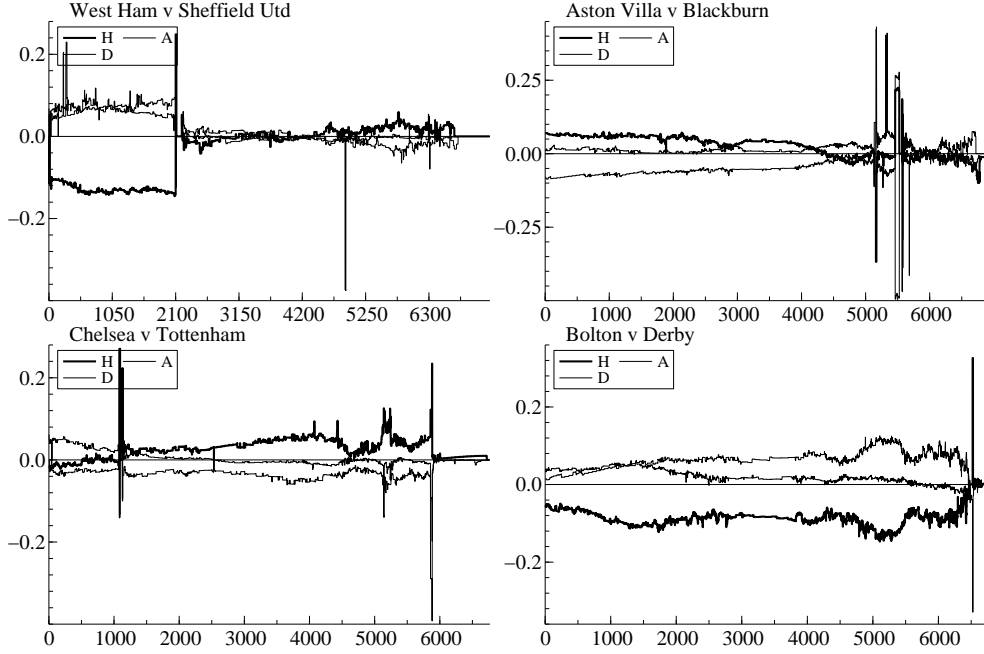
Figure 12: Illustrative plots of the difference between the Poisson-Multinomial minute-by-minute probability series and the Betfair implied probability series, i.e. $\widehat{P}_{Betfair,t} - \widehat{P}_{Poisson,t}$.

Market	Mean	Maximum	Standard Deviation
Home	1410.2	6047.1	1233.6
Away	1087.8	5655.7	1193.7
Draw	383.6	1601.1	310.7

Table 7: Sum of the squared deviations between the in-running probabilities predicted by our Poisson model and those implied by observed Betfair prices for all 547 English Premiership matches.

5 Liquidity and Efficiency

In this final section of the main paper, we explore the possibility that informational efficiency is related to the liquidity in individual betting markets. Looking across our full sample of matches, we observe considerable variation in liquidity per game. At one extreme, the Euro 2008 game between Italy and Romania saw over $50m traded in total. At the other end of the distribution, just $0.05m was bet on the UEFA Cup game between Marsaxlokk and Slaven Belupo. Figure 14 illustrates the variation in the sample.

In this section, we exploit this variation, together with a cleaner setting for efficiency tests, to offer a new perspective on the link between liquidity and informational efficiency of markets.

Whether and how liquidity affects a market's ability to aggregate information remains an interesting and open issue in the established finance literature (as well as in the developing literature on prediction markets, where its resolution has obviously crucial implications for market design).

A number of theoretical articles establish that liquidity is related to the informational efficiency of markets but three opposing views have emerged. In the first view, liquidity lowers the transaction cost (price impact) for informed arbitrageurs whose trades make prices more efficient. Thus, the
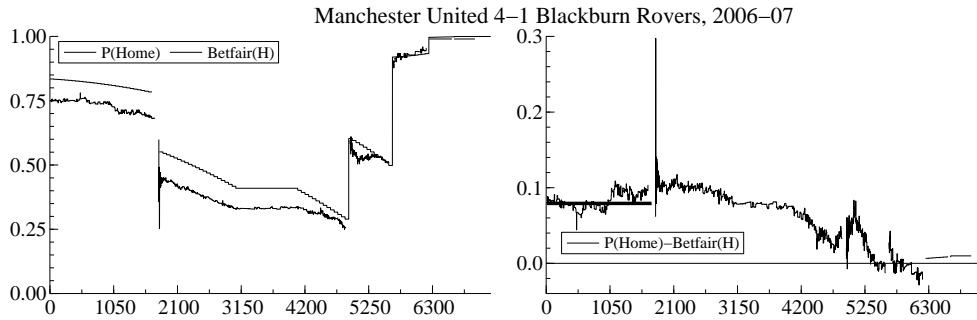
Figure 13: The fit between the Betfair implied probability series and our Poisson-Multinomial minute-by-minute probability series for Manchester United vs Blackburn Rovers (2006–07). The left panel depicts the two series (again, the smoother series is the Poisson series), and the right panel tracks the difference between the two.

Match	Home	Away	Draw
West Ham v Sheff Utd	1311.9	770.3	101.0
Aston Villa v Blackburn	1734.3	1160.5	133.4
Chelsea v Tottenham	183.1	22.8	93.9
Bolton v Derby	394.3	253.9	427.7
Man Utd v Blackburn	961.3	419.0	241.1

Table 8: Sum of the squared deviations between the in-running probabilities predicted by our Poisson model and those implied by observed Betfair prices for the matches depicted in Figures 11 and 13.

more liquid a market the more informative its prices. In many rational models, informed individuals will be more encouraged to trade aggressively when the market is liquid as they suffer reduced price impact. In addition, liquidity creates incentives to acquire information. A second and rival theoretical view is that liquidity represents non-informational trading (noise trading), so that we should expect more liquid markets to be less informationally efficient. In behavioral finance models, various factors may prevent rational agents from betting sufficiently aggressively against noise traders. For example, in De Long et al. (1990) rational traders may actually help propagate shocks caused by noise traders because they anticipate (rationally) that mispricing can worsen in the near term. Prices in liquid markets may actually be more inefficient as a result. Finally, it could be that liquidity and informational efficiency are in fact unrelated.

Empirical work has put these competing theories to the test. Many recent papers provide indirect empirical support for the first view, that securities mispricing is greater in illiquid markets (Kumar and Lee 2006; Sadka and Scherbina 2007; and Chordia, Roll, and Subrahmanyam 2008). However, Tetlock (2007) points out that none of these studies directly examines the deviation of securities prices from fundamental values because the terminal cash flows of the securities are unobservable. Instead, these analyses must rely on strong assumptions to get at the relationship between liquidity and mispricing. Tetlock's 2007 own empirical investigation utilizes data related to assets traded on an online prediction market. The advantage of his data set, he suggests, is that the assets traded have short lives and so reach their terminal values quickly. He finds "strong empirical support for the hypothesis that the prices of illiquid securities converge more
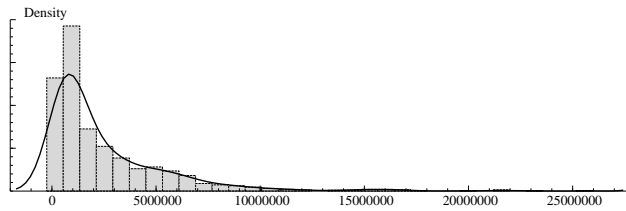
Figure 14: Distribution of sample games according to volume traded

quickly toward their terminal cash flows," so supporting the second strand of theoretical work. A limitation of Tetlock's investigation, from the perspective of semi-strong form efficiency, is that he does not observe event-relevant information.

Preliminary investigations based on our own dataset appear to support the view that there is no significant relationship between liquidity and informational efficiency in Betfair soccer markets, at least over the liquidity range we observe. Liquidity can be defined in a number of ways. In the current analysis, we use in-play trading volume as a liquidity proxy. Inserting a volume variable into our complete set of efficiency regressions we are unable to detect evidence of a link between volume traded and mispricing. Future work could deploy more sophisticated liquidity measures, such as the measure proposed by Martinez et al. (2001), and investigate also the relationship between variations in liquidity within a single game and the market's response to news arrival.

6   Concluding Remarks

The recent emergence of online betting exchanges has made it possible to obtain high-frequency data relating to bets placed "in-running" (during a live sports event). This implies a fertile new setting for empirical work, and in particular, it paves the way for a cleaner look at the topic of market efficiency. A market that is semi-strong form efficient updates swiftly and fully to publicly available information. A problem for those seeking to put this to the test in financial markets has been the possibility of news leakage not observed by the econometrician. In sports, however, major news (such as a goal in soccer) tends to break comparatively cleanly. We exploit this characteristic of sports events to offer a fresh study of efficiency. Prices for soccer-related markets are extracted from the live order book of the largest online betting exchange, Betfair.com, and tested for efficiency in relation to the arrival of goals. A complication particular to this exercise relates to the difficulty in determining whether any price drift following a goal should be interpreted as sluggishness in updating (and hence evidence of inefficiency), or be considered simply an efficient response to the passage of playing time (goalless periods of play being themselves price-relevant news). To overcome this identification issue, we exploit the naturally newsless half-time interval—we study matches where goals arrive on the cusp of the half-time break. Our findings suggest that prices swiftly and fully impound news.

A possible concern with our half-time identification might relate to the generalizability of our findings to news that arrives during minutes of play. The paper implements two further approaches to testing for efficiency in these markets, both of which provide a more direct perspective on the market's ability to incorporate information into prices whilst play is in progress. First, we exploit our large sample to conduct a meaningful test for drift in *average* prices during minutes of play. This test is a close relative of the calibration tests of prediction markets now often reported in that literature. In a semi-strong form efficient market, average post-goal prices should not drift during

32

minutes of play and we are unable to reject the null hypothesis of no such drift. Our conclusions are at odds with those from an earlier study based on a different sports betting data set (Gil and Levitt 2007 report some evidence of post-goal drift looking at a small set of Intrade soccer markets) and we suggest plausible reasons for this in the paper.

Our second approach to examining the efficiency of in-play updating to news exploits statistical knowledge of the underlying Poisson goal arrival process. Concretely, we compare the drift observed in Betfair markets to that which would arise under hypothetically efficient updating to the passage of time, where our estimation of this 'efficient' drift is based on updating according to a Poisson model fitted well to historical match data. Our tentative conclusions are that these Betfair markets behave largely 'as if' they are updating efficiently to the ticking down of the clock. There are some points of difference between the 'efficient' drift we create and the drift actually observed in Betfair prices and we discuss some limitations of the exercise in the paper. An interesting challenge for future work would be to fine tune our current model to allow for other sources of minor news that arrive in-play, such as the awarding of a free-kick or an injury. These are not currently captured in our econometric study but might form the basis for some updating (efficient or otherwise) on the part of bettors witnessing the game unfold. There may also be interesting interactions between the major news of a goal and the arrival of more minor in-play events.

Finally, our dataset features considerable variation in betting liquidity per game and we exploit this to contribute to an important and ongoing debate over the relationship between liquidity and informational efficiency. Including a liquidity proxy (total match volume) in our efficiency tests, we are unable to find support for any link between liquidity and mispricing.

References

Abarbanell, J. S. and V. L. Bernard (1992), 'Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior', *Journal of Finance* **47**, 1181–1207.

Asquith, P. (1983), 'Merger Bids, Uncertainty and Stock Holder Returns', *Journal of Financial Economics* **11**, 51–83.

Bachelier, L. (1900), Theorie de la Speculation, *in* P.Cootner, ed., 'The Random Character of Stock Market Prices', MIT Press, 1964, Cambridge, MA.

Bai, J. and P. Perron (2003), 'Computation and Analysis of Multiple Structural Change Models', *Journal of Applied Econometrics* **18**(1), 1–22.

Ball, R. and P. Brown (1968), 'An Empirical Evaluation of Accounting Income Numbers', *Journal of Accounting Research* **6**, 159–178.

Beaver, W. H. (1968), 'The Information Content of Annual Earnings Announcements', *Journal of Accounting Research* **6**, 67–92.

Chamberlain, E. (1948), 'An Experimental Imperfect Market', *Journal of Political Economy* **56**(2), 95–108.

Chan, L. K. C., N. Jegadeesh and J. Lakonishok (1996), 'Momentum Strategies', *Journal of Finance* **51**, 1681–1713.

Chan, W. S. (2003), 'Stock Price Reaction to News and No-News: Drift and Reversal after Headlines', *Journal of Financial Economics* **70**(2), 223–260.

Chen, N., R. Roll and S. Ross (2003), 'Economic Forces and the Stock Market', *Journal of Business* **59**(3), 383–403.

Chordia, T., R. Roll and A. Subrahmanyam (2008), 'Liquidity and Market Efficiency', *Journal of Financial Economics* **87**, 249–268.

Christiansen, J. (2007), 'Practical Experiments in Small Markets and Behaviours Observed', *The Journal of Prediction Markets* **1**, 17–41.

Daniel, K. D., D. Hirshleifer and A. Subrahmanyam (1998), 'Investor Psychology and Security Market Under- and Over-Reactions', *Journal of Finance* **53**, 1839—1886.

De Long, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann (1990), 'Positive Feedback Investment Strategies and Destabilizing Rational Speculation', *Journal of Finance* **45**, 379–395.

Doornik, J.A. (2007), *Object-Orientated Matrix Programming Using Ox*, 3 edn, Timberlake Consultants Press, London.

Doornik, J.A. and H. Hansen (2008), 'An Omnibus Test for Univariate and Multivariate Normality', *Oxford Bulletin of Economics and Statistics* **70**(s1), 927–939.

Fama, E. F. (1965), 'The Behaviour of Stock Market Prices', *The Journal of Business* **38**, 34–105.

Fama, E. F. (1970), 'Efficient Capital Markets: A Review of Theory and Empirical Work', *The Journal of Finance* **25**(2), 383–417.

Fama, E. F. (1998), 'Market Efficiency, Long-Term Returns and Behavioral Finance', *Journal of Financial Economics* **49**(3), 283–306.

Fama, E. F., L. Fisher, M. C. Jensen and R. Roll (1969), 'The Adjustment of Stock Prices to New Information', *International Economic Review* **10**, 1–21.

Gil, R. and S. Levitt (2007), 'Testing the Efficiency of Markets in the 2002 World Cup', *The Journal of Prediction Markets* **1**(3), 255–270.

Golec, J. and M. Tamarkin (1991), 'The Degree of Inefficiency in the Football Betting Market', *Journal of Financial Economics* **30**, 311–323.

Gray, P. K. and S. F. Gray (1997), 'Testing Market Efficiency: Evidence From The NFL Sports Betting Market', *The Journal of Finance* **52**(4), 311–323.

Hahn, R. W. and Tetlock P. C. (2005), 'Using Information Markets to Improve Public Decision Making', *Harvard Journal of Law and Public Policy* **29**, 214–289.

Hahn, R. W. and Tetlock P. C. (Eds.) (2006), *Information Markets: A New Way of Making Decisions*, AEI Press.

Hanson, R. (2002), Decision Markets, *in* A.Tabarrok, ed., 'Entrepreneurial Economics: Bright Ideas from the Dismal Science', Oxford University Press.

Hartzmark, S. and D. Solomon (2008), 'Efficiency and the Disposition Effect in NFL Prediction Markets', *Working Paper* .

Hendry, D., S. Johansen and C. Santos (2008), 'Automatic Selection of Indicators in a Fully Saturated Regression', *Computational Statistics* **23**(2), 337–339.

Im, K.S., M.H. Pesaran and Y. Shin (2003), 'Testing for Unit Roots in Heterogeneous Panels', *Journal of Econometrics* **115**(1), 53–74.

Jackson, A. and T. Johnson (2006), 'Unifying Underreaction Anomalies', *Journal of Business* **79**(1), 75–114.

Jarrell, G. and A. Poulson (1989), 'Stock Trading Before the Announcement of Tender Offers: Insider Trading or Market Anticipation?', *Journal of Law, Economics, and Organization* **5**(2), 225–248.

Karlis, D. and I. Ntzoufras (2003), 'Analysis of Sports Data By Using Bivariate Poisson Models', *The Statistician* **52**(3), 381–393.

Kothari, S. P. and J. B. Warner (1997), 'Measuring Long-Horizon Security Price Performance', *Journal of Financial Economics* **43**(3), 301–339.

Kumar, A. and C. Lee (2006), 'Retail Investor Sentiment and Return Comovements', *Journal of Finance* **61**, 2451–2486.

Laffey, D. (2005), 'Against the Odds—Case Study of Betfair.com', *The International Journal of Entrepreneurship Education* **3**(1), 99–114.

Le Roy, S. F. (1989), 'Efficient Capital Markets and Martingales', *Journal of Economic Literature* **27**(4), 1583–1621.

Levitt, S. and J. List (2007), 'What do Laboratory Experiments Measuring Social Preferences tell us about the Real World?', *Journal of Economic Perspectives* **21**(2), 153–174.

List, J. (2004), 'Testing Neoclassical Competitive Theory in Multilateral Decentralized Markets', *Journal of Political Economy* **112**(5), 1131–1156.

Maddala, GS and S. Wu (1999), 'A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test', *Oxford Bulletin of Economics and Statistics* **61**(S1), 631–652.

Manski, C. (2006), 'Interpreting the Predictions of Prediction Markets', *Economic Letters* **91**(3), 425–429.

Martinez, M. A., G. Rubio and M. Tapia (2001), 'Understanding the Ex-Ante Cost of Liquidity in the Limit Order Book: A Note', *Working Paper* .

McHale, I. and P. Scarf (2006), 'Forecasting International Soccer Match Results Using Bivariate Discrete Distributions', *Working Paper, Management and Management Sciences Research Institute, University of Salford* .

Ozgit, A. (2005), 'Auction versus Dealer Markets in Online Betting', *Working Paper, Department of Economics, UCLA* .

Patell, J. M. and M. A. Wolfson (1984), 'The Intraday Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements', *Journal of Financial Economics* **19**, 223–252.

Patterson, K. (2000), *An Introduction to Applied Econometrics: A Time Series Approach*, Palgrave Macmillan.

Plott, C. and S. Sunder (1988), 'Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets', *Econometrica* **56**(5), 1085–1118.

Sadka, R. and A. Scherbina (2007), 'Analyst Disagreement, Mispricing, and Liquidity', *Journal of Finance* **62**, 2367–2403.

Sunstein, C. (2006), *Infotopia: How Many Minds Produce Knowledge?*, Oxford University Press.

Tetlock, P. C. (2007), 'Does Liquidity Affect Securities Market Efficiency?', *Working Paper* .

Vaughan-Williams, L. (2005), *Information Efficiency in Financial and Betting Markets*, Cambridge University Press.

Waud, R. N. (1970), 'Public Interpretation of Discount Rate Changes: Evidence on the 'Announcement Effect', *Econometrica* **38**(2), 231–250.

Wolfers, J. and E. Zitzewitz (2004), 'Prediction Markets', *Journal of Economic Perspectives* **18**(2).

Wolfers, J. and E. Zitzewitz (2006), 'Interpreting Prediction Market Prices as Probabilities', *CEPR Discussion Paper* (5676).

Worrell, D. L., W. N. Davidson III and V. M. Sharma (1970), 'Layoff Announcements and Stockholder Wealth', *Academy of Management Journal* **34**(3), 662–678.

7 Appendix: Supplementary Material

Table A.3: Summary of betfair trading for the $1,206$ soccer matches in our sample

Competition	Matches	Volume of Bets Matched ($)					
		Total	At Kick-off	% in-match	Per Sec	HT Per Sec	Cusp HT Per Sec
Averages							
Premiership 2006–07	250	5,378,973	2,714,744	44	429	272	503
Premiership 2007–08	235	6,818,284	3,567,931	46	516	474	1,333
Premiership 2008–09	62	7,824,675	3,998,445	47	607	296	366
Champions League	165	5,275,070	2,940,899	45	371	257	423
UEFA Cup	249	2,171,738	1,009,800	54	184	136	244
Intertoto Cup	14	715,631	279,787	63	69	36	71
International	42	2,003,677	949,702	60	167	103	209
Scottish Premier	64	1,483,680	611,827	58	138	89	99
Asian Cup	24	1,004,772	465,570	55	599	49	67
Euro 2008	25	31,104,642	15,142,316	52	2,534	1,586	2,373
Euro 2008 Qualifiers	76	3,656,337	1,911,386	46	277	201	262
Total	**1206**						
Average(average)		**6,164,540**	**3,089,934**	**51**	**527**	**319**	**552**
Maxima							
Premiership 2006–07		24,024,000	12,103,000	76	1,892	1,726	1,726
Premiership 2007–08		22,496,000	14,565,600	75	1,846	12,613	12,613
Premiership 2008–09		23,685,400	14,318,680	78	2,033	1,003	766
Champions League		34,686,000	17,560,240	75	2,718	2,896	2,896
UEFA Cup		13,711,080	6,576,800	90	1,132	2,193	1,321
Intertoto Cup		1,493,980	791,722	77	149	116	116
International		12,819,260	7,449,440	93	852	602	602
Scottish Premier		6,186,220	2,479,120	82	588	1,043	183
Asian Cup		1,900,338	975,706	77	1,129	96	96
Euro 2008		51,555,000	32,661,600	68	3,503	2,817	2,373
Euro 2008 Qualifiers		29,957,400	12,162,720	85	2,825	1,794	814
Maximum		**51,555,000**	**32,661,600**	**93**	**3,503**	**12,613**	**12,613**
Minima							
Premiership 2006–07		366,180	212,860	8	16	1	14
Premiership 2007–08		302,200	149,036	16	24	14	90
Premiership 2008–09		1,090,156	544,394	20	56	19	88
Champions League		340,364	126,560	12	12	3	5
UEFA Cup		54,548	21,281	17	5	0	13
Intertoto Cup		241,854	69,976	30	23	20	27
International		103,343	7,367	21	15	3	37
Scottish Premier		242,090	79,294	21	15	4	41
Asian Cup		516,780	166,147	36	273	6	16
Euro 2008		7,914,060	4,242,060	32	583	350	2,373
Euro 2008 Qualifiers		486,576	130,456	6	6	1	4
Minimum		**54,548**	**7,367**	**6**	**5**	**0**	**4**

| Variable | Home | | | Away | | | Draw | | | Overall | | |
	Normality	Skewness	Kurtosis	Normality	Skewness	Kurtosis	Normality	Skewness	Kurtosis	Normality	Skewness	Kurtosis
$pr_{t,c,m}$	99.4%	79.4%	98.9%	99.2%	79.3%	98.2%	98.9%	78.1%	97.7%	99.2%	78.9%	98.3 %
$\widehat{\varepsilon}_{t,c,m}$	100.0%	98.9%	99.5%	100.0%	99.1%	99.9%	100.0%	98.3%	99.2%	100.0%	98.8%	99.6%

Table A.4: Normality tests for both price series and residuals. Test is evaluated at 5% significance level, and the proportion of our 1206 matches where the null hypothesis of Normality, normal skewness or normal kurtosis is reported.

7.1   Half-time identification: Monte Carlo Simulations

Because of the non-standard null hypothesis for unit root testing, and because our test equation for our half-time testing also contains an extra lag and constant term, all of which must be restricted to zero when testing for a random walk without drift, we simulate critical values. An additional feature of our data is that, because of its high-frequency nature, the residual distributions tend to be more leptokurtic than standard normally distributed variables, and hence we simulate generating our error terms from a t-distribution with three degrees of freedom to capture this feature of our data.

In order to provide assurance that the leptokurtic assumption is appropriate, we provide in Table A.4 the results of the Doornik and Hansen (2008) test for normality. In the first row of the table, we provide results for the actual price series from each match, and in the second row we provide the results for the residuals of our AR(2) test regression model. What is reported in each row is the proportion of tests that reject the null hypothesis when the test is evaluated at the 5% significance level. The Doornik and Hansen test is the sum of two individual test statistics for skewness and kurtosis, each of which is χ_1^2 distributed. Hence we also provide in each row the outcomes of these individual tests. The overwhelming result from this is that the assumption of normality, and of normal kurtosis, is rejected: The rejection frequencies range from 97.7% to 100%.

The impact of this leptokurtic error distribution is that the simulated 1% critical value is substantially higher than it would be if our errors were generated using a standard normal distribution.

We generate our critical values by simulating under the null hypothesis of semi-strong form market efficiency and generating N time series of T observations, and calculating the test statistics for our tests of interest for each replication. When considering the M-W test we simulate the time-series ADF test 100,000 times first in order to generate the small-sample distribution from which to extract the p-values to be used in the panel test. We implement 1000 replications, and the simulations were written in the Ox programming language (Doornik, 2007).
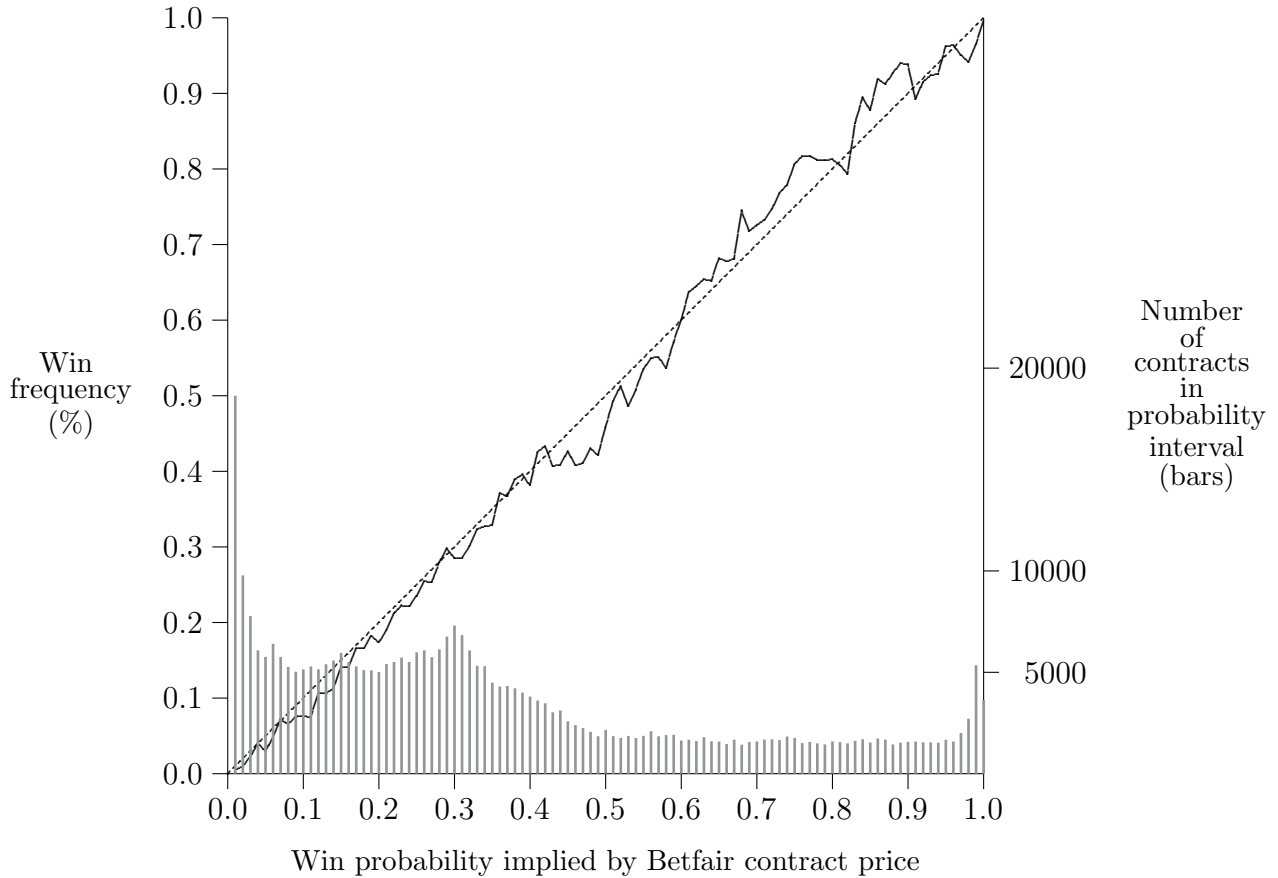
Figure 15: In-running calibration test: win probabilities implied by betfair prices plotted against observed win frequencies. This plot is based on all observations in our sample of $1,206$ matches. The implied win probability for each Betfair contract in each match is computed using the volume-weighted average Betfair prices for each minute of the match. We plot these implied win probabilities against their win frequencies in the sample and look to assess the fit of this series to the 45° line; it is impossible to tell whether a single contract priced to imply probability X% was priced efficiently, but in an informationally efficient market a large sample of contracts priced at X% will win X% of the time, implying a good fit to the 45° line. Our plot reveals a strong fit to this line. The bars in the plot indicate the number of observations in our sample that fall into each probability bracket. For example, tracking minute-by-minute prices for all 1,206 matches during minutes of play, we observe around $20,000$ contracts whose prices imply a win probability in the bracket 0–0.01. Most of these observations will be associated with later stages of games, where often considerable evidence against at least one of the three outcomes has accumulated. Some previous studies have reported evidence of a so-called 'favourite-longshot' bias in sports betting markets—a tendency for underdogs (prospects associated with low probabilities) to be over-priced and strong favourites (outcomes associated with high probabilities) to be under-priced (see Vaughan-Williams 2005 for an overview of research on this bias). We are unable to detect this bias in our data based on this plot and other investigations conducted so far. It could be that a calibration analysis that is more disaggregated along the time dimension (e.g., which looks separately at the calibration of betfair prices during earlier minutes of matches) might reveal traces this or other biases, possibly showing a tendency for these to 'wash out' as time elapses and outcome-relevant information accrues. This is a current line of investigation.