Institute for Strategy and Business Economics
University of Zurich

Working Paper Series
ISSN 1660-1157

Working Paper No. 120

The Validity of Models on the Information Content of Trades

Leif Brandes, Egon Franck and Erwin Verbeek

August 2009

The Validity of Models on the Information Content of Trades

Leif Brandes*, Egon Franck, Erwin Verbeek

(version: August 2009)

Abstract

Measuring the (private) information content of stock trades is an important topic in market microstructure research. A common problem of stock markets is that it is ex ante not possible to separate phases where the scope for asymmetric information is likely to be broader from those, where there is less exposure to traders with superior information. Such a distinction is needed to directly test the reliability of the proposed measures.
This paper applies a unique data set from the betting market to provide such a direct test. We exploit the fact that betting occurs preplay, i.e. before match start, as well as inplay. We test the hypothesis that measures of private information will be less pronounced for inplay betting quotations (where all actions on the court are publicly observable, and trading on private information is unlikely) than for preplay betting quotations. Based on more than 23,000 transactions for the 2008 Wimbledon final, and five additional matches, we present first empirical support for this relation.

* Corresponding author: University of Zurich, Institute for Strategy and Business Economics, Plattenstrasse 14, 8032 Zurich, email: leif.brandes@isu.uzh.ch, phone: +41 44 634 29 61

Introduction

Several measures on the extent of information asymmetry in financial markets have been developed in the literature on market microstructure. Researchers generally accept the notion that trades convey information in the presence of heterogeneously informed agents. It follows that agents will update their conditional expectation of a security's (terminal) value based on observed trading activity in the market, resulting in a persistent impact on the asset's price. Following Hasbrouck (1991a), a certain consensus has evolved by which „the magnitude of the price effect for a given trade size is generally held to be a positive function of the proportion of potentially informed traders in the population, the probability that such a trader is infact informed, and the precision of the private information" (p. 179). Thus, the price impact triggered by (unexpected) trades serves as a common proxy for the extent of trading activity based on superior information (e.g., Glosten & Harris, 1988; Hasbrouck, 1991a, 1991b; Huang & Stoll, 1997 and Madhavan, Richardson & Roomans, 1997).

But how well do these measures perform in terms of their reliability and validity? For example, do previous measures indeed find larger price impacts in market phases which are known to be characterized by, e.g., a higher proportion of potentially informed traders in the population? If one could unambiguously differentiate between markets or trading periods with respect to their extent of information asymmetry, this would serve as a direct test for the reliability of such models.

This paper provides such a test for several[1] models on the information content of security trades, which, to our knowledge, has never been performed before. This gap in the literature can mainly be attributed to the lack of appropriate data sets in traditional financial markets. In stock markets, for example, it is generally impossible to provide *ex-ante* clear-cut break points which separate market phases of low scope for information asymmetries from periods that expose agents to greater risk of superior information.

[1] Within this version, we focus on the model by Hasbrouck (1991a).

To circumvent this problem, we depart from previous estimation approaches and analyze a novel data set from the online betting market *Betfair*. This market has increasingly been attracting researchers' interest within recent years as it has become comparable (in terms of economic importance) to more traditional financial markets. Croxson & Reade (2008), for example, state that "*Betfair* processes around two million trades a day - six times the number of trades on the London Stock Exchange." Our data comprises full transaction records for more than 100 tennis matches at the 2008 Wimbledon Championship. This tournament is generally held to be the most prestigious of all tennis championships, and is characterized by large trading volumes and high media interest.

For our study, however, the key feature of this data lies in the fact that betting on sports events happens in continuous time before and after the start of the event. This allows us to separate 'inplay' betting activity from 'preplay' bets. Tennis matches are particularly suited in this respect, as matches frequently exhibit durations of more than three hours - this guarantees a longer inplay transaction phase than many other sports disciplines, e.g., soccer. We suggest that these two phases structurally differ in the way that relevant information diffuses into the market. Within our empirical strategy we identify 'inplay' ('preplay') phases as those where the proportion of bettors possessing superior information should be lower ('higher'). The reasoning behind this is that in the inplay period virtually all relevant information is produced on the court and published in real time through television live coverage[2]. Thus, the market participants in the inplay phase can be seen as endowed with homogeneous information, if we assume that most inplay-traders are actually watching the match (or, at least, following it closely on the web). Therefore, inplay trading activity should only mirror different beliefs of investors (e.g., because a certain player performance can be ambiguously interpreted, such as an early break) but should not be caused by the risk of superior, i.e., asymmetric information. In comparison to that, private information may well play a role *before* match start. Here, relevant information on the outcome of the match can potentially be produced from geographically dispersed places and,

[2] The picture of central information provision in the inplay period is further reflected in the fact that *Betfair* provides a live ticker on the inplay results.

even more important, it is published through different media channels at different points in time. Within this paper, an investor/ bettor is assumed to possess private information in the following sense. First, a bettor may have access to information which is publicly observable but of which some bettors are not aware yet. Consider, for example, an individual who closely follows the news on a webpage such as *CNN.com* (USA) or *The Sun* (UK). For such pages it is nowadays common to provide a ticker on the latest news[3]. By the time, breaking news has been published, this bettor may be able to exploit existing limit orders from other bettors (the order book is public at *Betfair*, see also below) who have not come across this news yet. This gives rise to the temporary existence of *de facto* private information in the market: not all information is shared by all bettors at all times[4]. Second, the bettor may have real insider knowledge as classically understood in the financial economics literature. To justify why some bettors may have insider information in sports markets, consider the example of Scottish tennis professional Andy Murray. As of June 2009, this player had a team of five permanent coaches, psychologists and physicians[5]. If one reckons that these team members, as well as Murray and families, may know about health problems, strategy and/or new training methods of Murray earlier than the public, there arises a chance for the exploitation of inside information before match begin. For each match, the number of potential insiders can thus be assumed to lie between 10-20 individuals (5-10 per player). Due to this limited number of individuals, we expect temporary private knowledge to be the major source of private information asymmetries affecting investors' betting strategies, but state that the classically defined concept of insider information may actually also be encountered in this type of market.

---

[3] At *www.thesun.co.uk* the time since publication is shown, as well. News from all areas (e.g., sports, TV celebrities) is published in real-time.

[4] Support for this idea comes from the literature on limited investor attention. See, for example, Cohen & Frazzini (2008) and the references therein.

[5] Among the *ATP* Top 10 players, this number lies between 3 and 5. These figures are very conservative, as they only refer to permanent team members. Based on information from internet fan forums, we got the impression that many players hire additional team members for specific occasions.

Based on a selection of six matches[6], we thus test the hypothesis that the price impact from unexpected trades is larger for preplay transaction than for inplay transactions. We apply the vector autoregressive approach proposed by Hasbrouck (1991a) which models a security's price change as a function of past price changes and past signed trades (+1 for buys and -1 for sells). We show that it is straightforward to operationalize this model in a betting environment. A match's fundamental value is identified by the winning probability of player i relative to the winning probability of player j. This incorporates the notion that prices in prediction markets reflect the market's expectation on the likelihood of each outcome[7].

By choosing tennis matches with two possible outcomes (player i to win the match or player j to win the match) rather than soccer matches with three possible outcomes (home win, away win, draw), we obtain a situation in which betting on player i is conceptually identical[8] to betting against player j (which does not hold for matches offering the chance of a draw). This allows for a straightforward derivation of signed trades and - subsequently, signed trading volumes.

Our empirical results largely support the appropriateness of the model by Hasbrouck (1991a). When using signed trades as an explanatory variable for price changes, we find support for our hypothesis in four out of six cases. Noteworthy, the 2008 Final, which exhibited the largest trading volume (USD 100 Mio), and was the most frequently traded in our sample, belongs to these four matches. Here, an additional preplay buy for player i in $t = 0$ results in a persistent increase in his relative winning probability of 0.0001 ($\alpha = 0.05$). The impact of inplay buys, however, is not statistically significant. As all signs for the estimated coefficients in the VAR model for this match are in line with the documented findings in Hasbrouck (1991a), we are confident about the appropriateness of our estimation approach. For two out of six

[6] These matches include the final, one semi-final, one quarter-final, and three matches from previous rounds in order to check the robustness of the results across the development of the tournament.

[7] We will discuss the appropriateness of this assumption in the next section.

[8] Here, 'conceptually identical' means that a bettor will in both cases be long in player i's performance. However, as we discuss in section 5.4.1, there are some procedural differences in the placement of bets on a player, called backing, and bets against a player, called laying.

matches, however, inplay price reactions are larger than preplay price impacts.

Subsequently, we estimate VAR models based on signed transaction volumes for the three matches exhibiting the largest values on total volume matched. For one match, again the Final, we find the hypothesized result. One match, however, shows a larger price effect from inplay transactions than from preplay transactions and for the final match, the inplay model suggests a lag order of 0.

The remainder of the paper is organized as follows. The next section shortly reviews previous attempts in the literature to measure the informational content in financial transactions. Then, the model by Hasbrouck (1991a) is explained in detail. The fourth section presents our data and estimation approach. Subsequently, the corresponding empirical results are presented and discussed. The final section concludes.

Related literature

Many studies have modeled transaction and price dynamics in markets where asymmetrically informed agents submit buy and sell orders against preexisting bid and ask quotations from a market maker[9]. „In these theoretical approaches, the underlying (and generally unobservable) characteristics of the informational asymmetries are prevalence, precision and timeliness of the private signals, and the extent of competition among informed traders. Agents' beliefs about these characteristics determine two observable features of the market: the bid-ask spread, and the impact of a trade on the security price" (Hasbrouck, 1991b, p. 572). Both features have been exploited in empirical research, which is why empirical studies can be grouped by their use of one these features. Analyzing bid-ask spreads under asymmetric information has been the focus of studies by McInish & Wood (1992), Chiang & Venkatesh (1988), and Foster & Viswanathan (1993, 1995). As mentioned by Hasbrouck (2007), this approach may be useful if the econometrician only observes in-

[9] See Bagehot (1971), Copeland & Galai (1983), Glosten & Milgrom (1985), Kyle (1985), Easley & O'Hara (1987), Glosten & Harris (1988), Glosten (1987,1989), Admati & Pfleiderer (1988), Foster & Viswanathan (1990) and Back (1992). For a general introduction to the theory of market microstructure, see O'Hara (1995).

formation on bid-ask spreads. However, once information on trades and trade sizes is feasible, the second branch of studies should have superior explanatory power. Among such studies are those by Glosten & Harris (1988), Stoll (1989), Foster & Viswanathan (1993, 1995), and Hasbrouck (1988, 1991a). It is to this later study's modeling approach that we will rely on in this article.
Before turning to this model, however, we shortly need to point out that there has evolved a third approach in more recent years which develops and implementrs methods to measure information asymmetry by focusing on the trade (signed order flow) process. The summary proxy for asymmetric information in this models is referred to as PIN (probability of informed trading). Studies in this direction include Easley, Kiefer & O'Hara (1996, 1997), Easley, Kiefer, O'Hara & Paperman (1996), Easley, Hvidkjaer & O'Hara (2002), and Easley, Engle, O'Hara & Wu (2008).

The Hasbrouck model

The model by Hasbrouck (1991a) is based on the following framework. He considers a security's price process in the presence of a market maker who posts bid and ask quotes which are denoted by q_t^b and q_t^b, respectively.
Regarding the midpoint of the quotes - the key variable in the model, it is assumed that

$$E\left[\left(q_s^b - q_s^a\right)/2 - Z_\tau | \Phi_t\right] \to 0 \tag{1}$$

as $s \to \tau$. Here, Z_τ denotes the final value of the security at some convenient terminal date τ in the future and Φ_t relates to the set of public information at time t. In words, Equation (1) states that, given today's information, the midpoint of the quotes eventually, i.e., for s getting closer to the terminal date τ, reflects the securities true fundamental value. Put differently, bid and ask quotes will eventually be symmetric around the conditional expectation. The important aspect of this Equation is that it allows for transitory deviations of the midpoint from the fundamental value of the security.
Let $r_t = \left(q_t^b + q_t^a\right)/2 - \left(q_{t-1}^b - q_{t-1}^a\right)/2$ denote the change in the midpoint of the quotes between time $t-1$ and t. The timing convention of this model states

that trades occur before the revision of trades. Transactions are characterized by their signed volume and are denoted by x_t. Here, the convention is followed that transactions above (below) the midpoint are identified as buys (sells). Volumes are positive if the underlying trade is a buy and negative for sells.

Having observed x_t, the market maker then posts new bid and ask quotes. In the absence of transaction costs, and if x_t denotes the only innovation in public information at time t, r_t fully reflects the information content of x_t.

Regarding the dynamic evolvement of r_t, several models on market microstructure show that price changes may exhibit a potential for autocorrelation: „Price dicreteness, for example, may induce threshold effects, since quote revisions may not be optimal before a series of trades of the same direction has occurred. [...] Inventory control effects, lagged adjustment to information and exchange-mandated price smoothing also involve serial dependence“ (Hasbrouck, 1991a, p. 183).

This suggests to specify

$$r_t = a_1 r_{t-1} + a_2 r_{t-2} + \ldots + b_1 x_t + b_2 x_{t-1} + \ldots + v_{1,t} \ . \tag{2}$$

In words, Equation (2) states that the revision of the midquote is a function of past midquote revisions, past transactions and a price innovation $v_{1,t}$.

As the very same mechanisms mentioned above will also affect transaction dynamics, Hasbrouck (1991a) models

$$x_t = c_1 r_{t-1} + c_2 r_{t-2} + \ldots + d_1 x_{t-1} + d_2 x_{t-2} + \ldots + v_{2,t} \ , \tag{3}$$

where $v_{2,t}$ reflects unanticipated (innovative) trade components. A key insight is the notion that if there is to be any information impact from trades on prices, it has to come from this term (the remainder of the left hand side denotes the expectation of trades based on the linear projection of trades on past price changes and the history of trades).

The central contribution by Hasbrouck (1991a) lies in modeling (2) and (3) as a vector autoregressive (VAR) model (see e.g., Hamilton (1994)). It is as-

sumed that both disturbance terms have zero means and are serially and jointly uncorrelated.

An important distinguishing feature of the information impact of trades lies in its persistency. Therefore, Hasbrouck (1991a) measures the cumulative m-step price impact of a $t = 0$ trade innovation of size $v_{2,0}$ by

$$\alpha_m(v_{2,0}) = \sum_{t=0}^{m} E[r_t | v_{2,0}] .$$

Note that (1) implies

$$\alpha_m(v_{2,0}) \longrightarrow E[Z_r | v_{2,0}] - E[Z_r]$$

for m increasing. In other words, „the expected cumulative quote revision converges to the revision in the efficient price. For this reason, $\alpha_m(v_{2,0})$ possesses an interpretation as the information revealed by the trade innovation [...], the VAR modeling strategy applied to trades and quotes allows, in principle at least, a resolution between private information (the trade innovation) and public information (the quote revision innovation)“ (Hasbrouck, 1991a, p. 185).[10]

Empirical framework

The data

We collected detailed betting market information from the online person-to-person betting platform *Betfair*. Here, individuals contract their opposing opinions with each other. On the online platform, they can post the prices under which they are willing to place a bet - on or against - a given event. The latent demand for wagers is collected and presented in the limit order book. It displays the most attractive odds with the corresponding available

[10] The reader should note that this resolution may in practice be much more ambiguous than in practice. To see how this would alter the interpretation of $\alpha_m(v_{2,0})$ as a measure of private information, the reader is referred to the careful discussion by Hasbrouck (1991a).

volume in a canonical manner and the book is public knowledge. The bettor has the choice to either submit a limit order and wait for another participant to match his bet, or to submit a market order and directly match an already offered bet. As a result, there is a continuous double auction process taking place at the online platform. If two bettors with opposing opinions agree on a price, their demands are automatically translated into a transaction. After the bets have been matched, both individuals hold a contract on a future cash flow. The size of the cash flow is determined by the price of the contract, and the direction of the cash flow is tied to the outcome of the underlying event. The provider of the platform charges a commission fee on the bettors' net profits.

Let us shortly comment on the betting procedure at *Betfair* through an example for a tennis match between *Andy Murray* and *Mardy Fish*.[11] Imagine, you believe *Andy Murray* to win this match and you would like to bet on this outcome. Note that there are two ways for you to do this; you could either back *Murray* or lay *Fish*. In particular, this means that you could immediately stake up to 420 GBP at odds 1.17 to back *Murray*, and up to additional 26,754 GBP at somewhat less attractive odds of 1.16. Note that *Betfair* uses decimal odds which are inclusive of stake. In other words, a 10 GBP bet to back *Murray* at odds of 1.17 would result in a gross return of GBP 11.70 (1.70 GBP profit plus 10 GBP stake). Alternatively, you could 'lay' *Fish*. Here, it is important to understand that all odds are displayed from a backer's perspective. Thus, a quote of 7 and 156 GBP implies that someone (or some combination of users) has submitted limit-orders hoping to back *Fish* asking for odds of 7 (i.e., slightly better than the prevailing market odds). In case you decide to lay *Fish*, this would mean that your acceptance of a 10 GBP bet corresponds to risking 70 GBP for a win of 10 GBP.

Originally, our sample comprised second-by-second information on all markets for 106 tennis matches during the 2008 Wimbledon Championship Tournament. The advantage of choosing tennis matches (e.g., in comparison to soccer matches) lies in the fact that laying *Fish* and backing *Murray* can both be interpreted as 'buying the *Murray* asset'.

[11] The following discussion, as well as the example on the betting procedure closely follows the description in Croxson & Reade (2008).

The choice of the Wimbledon tournament was based on the notion that it belongs to the most traditional and prestigious tournaments in professional tennis. Its significance is further reflected in being part of the Grand Slam series and its high degree of media coverage.

For all matches within our sample, our data comprises detailed information on (i) each player's three best back prices (including the prevailing available volumes) (ii) each player's three best lay prices (including the prevailing available volumes) (iii) the last price matched (including the associated volume) for each player. These data are available for any given second in the preplay period (between two to three hours prior the start of the match) as well as in the inplay period (again, about two to three hours).

Table 1 displays descriptive statistics on the matched volume (total & preplay), number of transactions (total and preplay) as well as the total number of observations in our sample.

Table 5.1
Trading volumes and number of transactions per match for the Wimbledon 2008 tournament

	mean	st. dev.	min.	max.	nr. of obs.
volume per match (total)	8,189,153	1.30e+07	88,971	4.91e+07	311,225
volume per match (preplay)	1,359,040	1,870,187	25,140	6,292,035	58,625
transactions per match (total)	6,249	6,072	375	23,911	311,225
transactions per match (preplay)	1,329	1,306	11	4,566	58,625

Notes: The table presents some summary statistics on trading volumes (in GBP) and number of transactions per match for the Wimbledon 2008 tournament.

The Table reveals the economically significant size of volume turnover and number of transactions for the matches within our sample. For example, the average matched volume per match lay around 16 Mio USD and the average number of trades per match equals 6,250. Within the 2008 Final, the volume approximated the remarkable value of 100 Mio USD, out of which 12 Mio had been traded in the preplay trading window. The match was characterized by vivid trading activity before match start, showing 4566 transactions within a

duration of merely three and a half hours, i.e. one transaction per 2.75 seconds. The average volume of these transactions amounted to 1,675 USD.

Estimation approach

As discussed in section 5.3, the model by Hasbrouck (1991a) relies on quote midprices as a proxy for the security's fundamental value. To operationalize this approach in a betting market's environment, we first need to decide on the fundamental value's equivalent. We choose the relative winning probability of player i: Recall that at any point in time t, we have two observations for the match between player i and player j: Back and Lay odds on player i, as well as back and lay odds on player j. Based on this information, we derive mid-price odds for each player k, $k \in \{i, j\}$ for every time t, $t = 1, ..T$:

$$P_k^t = \frac{[bp_k^t + lp_k^t]}{2} \tag{4}$$

Where bp_k^t and lp_k^t denote the best available back and lay quotes for player k at time t, respectively. Obviously, player i's series of quotations cannot be treated in isolation as information on player j will also affect the winning probability of player i.

An approach, which takes this interdependence into account is the formulation of player i's relative winning probability, which is defined by

$$P_{ij}^t = \frac{P_i^t}{P_j^t} \tag{5}$$

which can be interpreted as the match's fundamental value.

Price changes are than defined by

$$dP_{ij}^t = \frac{P_i^t + dP_i^t}{P_i^t + dP_j^t} - \frac{P_i^t}{P_j^t} \tag{6}$$

where $dP_k^t = P_k^t - P_k^{t-1}$. This approach reflects the possibility that there may be no change in the quotation of player i, but only in the quotation of player j which would still alter the relative winning probability of player i.

To ease the exposition of the results for the reader, we transform (5) and (6) into relative probabilities, which are derived[12] from the inverse of a player's quotation. Therefore, all estimation results will be based on the changes in the relative winning probabilty of player i:

$$dprob_{ij}^t = \frac{\frac{1}{P_i^t + dP_i^t}}{\frac{1}{P_j^t + dP_j^t}} - \frac{\frac{1}{P_i^t}}{\frac{1}{P_j^t}} = \frac{P_j^t + dP_j^t}{P_i^t + dP_i^t} - \frac{P_j^t}{P_i^t}. \tag{7}$$

The second ingredient in the model by Hasbrouck (1991a) are signed trades, where buys are coded as (+1) and sells are coded by (-1). As there are two players for each match, each having back and lay quotations, there are two ways for a bettor to 'buy' player i, i.e. to be long in the underlying player's performance: The bettor can either back player i or lay player j - in both cases, the bettor profits from a win of player i. An analogous reasoning applies for selling decisions.

Taking player i as the underlying of this contract, we derive the sign of the transaction as follows:

$$x_{ijt}^0 = \begin{cases} +1\text{: } lpm_i^t < P_i^{t-1} \ \ or \ \ lpm_j^t > P_j^{t-1} \\ -1\text{: } lpm_i^t > P_i^{t-1} \ \ or \ \ lpm_j^t < P_j^{t-1} \end{cases} \tag{8}$$

Where lpm_k^t denotes the last price matched for player k before time t. Equation (8) thus reflects the above mentioned strategies to buy and sell the underlying asset.

[12] There is currently a debate in the literature on prediction markets, whether implied probabilities can indeed be inferred from this approach as questioned by Manski (2004).

Based on $dprob_{ij}^{t}$ and x_{ijt}^{0} we specify the same kind of VAR model[13] as Hasbrouck (1991a). We test our hypothesis by separating preplay observations from those that occurred inplay leading to

$$\begin{aligned} dprob_{ij}^{t} &= \sum_{m=1}^{M_1} a_m dprob_{ij}^{t-m} + \sum_{m=0}^{M_1} b_m x_{ij}^{t-m} + v_1^t \\ x_{ij}^{t} &= \sum_{m=1}^{M_1} c_m dprob_{ij}^{t-m} + \sum_{m=1}^{M_1} d_m x_{ij}^{t-m} + v_2^t \end{aligned} \tag{9}$$

for preplay quotations and to

$$\begin{aligned} dprob_{ij}^{\tau} &= \sum_{m=1}^{M_2} a_m dprob_{ij}^{\tau-m} + \sum_{m=0}^{M_2} b_m x_{ij}^{\tau-m} + v_1^\tau \\ x_{ij}^{\tau} &= \sum_{m=1}^{M_2} c_m dprob_{ij}^{\tau-m} + \sum_{m=1}^{M_2} d_m x_{ij}^{\tau-m} + v_2^\tau \end{aligned} \tag{10}$$

for inplay observations. The reader should note that we denote preplay transaction time by t and inplay transaction time by τ to avoid misunderstandings. This convention will be followed throughout the remainder of this paper. We also emphasize that we allow the number of lags to differ between preplay and inplay observations. To determine the relevant number of lags for each

[13] Performing augmented Dickey-Fuller (ADF) and KPSS tests on the stationarity of $dprob_{ij}^{t}$ and x_{ij}^{t}, the null hypothesis of a unit-root is always rejected for the ADF. For the KPSS the null of stationarity is not rejected for the $dprob_{ij}^{t}$ series, but for the x_{ij}^{t} series. However, due to the generally low power of unit root tests, and the fact that the x_{ij}^{t} series is bounded below (-1) and (+1) by construction, we decide to estimate standard vector autoregressive models instead of vector error correction models.

model, we rely on the Schwarz-Buchanan information criterion (SBIC) which is known to yield a consistent estimate of the unknown lag length[14].

Based on the estimation approach and the theoretical considerations, we formulate our hypothesis as follows:

For each match, and for $m \geq m^*$ *we expect* $\alpha_m(v_2^\tau) \leq \alpha_m(v_2^t)$. *In other words, the persistent price effect is weakly larger for preplay private information innovations than for inplay innovations.*

Empirical results

To economize on space, we only give full estimation results for the 2008 Wimbledon Final. For selected additional matches, only graphical illustrations of the cumulative price impact are displayed.

Signed trades

We follow the line of exposition in Hasbrouck (1991a) and discuss empirical evidence on the bivariate model (9) and (10) first. Estimation results on the 2008 Wimbledon final are displayed in Table 2. An important feature of this model is the varying lag length for preplay and inplay transactions. Throughout our discussion, the coefficients on $x_t, \ldots, x_{t-M}$ in r_t Equation (2) will be of greatest interest. We begin our discussion with preplay transactions.

[14] See e.g., Lütkepohl (2006). SBIC values have been obtained from the *varsoc* command in STATA. The maximum number of lags to be included was always set to 30.

Table 5.2
Estimates of the bivariate vector autoregressive model for 2008 Wimbledon final

preplay						inplay					
dprob			*x*			*dprob*			*x*		
a_1	-0.079	-5.34	c_1	-64.029	-9.00	a_1	0.039	5.39	c_1	-4.874	-15.77
a_2	-0.051	-3.40	c_2	-24.771	-3.45	a_2	0.021	2.86	c_2	-3.154	-10.12
a_3	-	-	c_3	-	-	a_3	0.048	6.55	c_3	-2.664	-8.52
a_4	-	-	c_4	-	-	a_4	0.022	2.89	c_4	-2.036	-6.49
a_5	-	-	c_5	-	-	a_5	0.007	1.00	c_5	-1.661	-5.29
a_6	-	-	c_6	-	-	a_6	0.006	0.79	c_6	-1.904	-6.06
a_7	-	-	c_7	-	-	a_7	-0.018	-2.45	c_7	-1.374	-4.37
a_8	-	-	c_8	-	-	a_8	0.010	1.36	c_8	-1.504	-4.78
a_9	-	-	c_9	-	-	a_9	-0.013	-1.77	c_9	-1.821	-5.79
b_1	0.000	1.22	d_1	0.359	23.60	b_1	0.000	0.47	d_1	0.253	34.53
b_2	0.000	2.03	d_2	0.084	5.70	b_2	0.000	-0.79	d_2	0.083	10.97
b_3	-	-	d_3	-	-	b_3	0.000	-0.41	d_3	0.055	7.21
b_4	-	-	d_4	-	-	b_4	0.000	0.33	d_4	0.032	4.26
b_5	-	-	d_5	-	-	b_5	0.000	0.21	d_5	0.050	6.64
b_6	-	-	d_6	-	-	b_6	0.000	1.47	d_6	0.021	2.72
b_7	-	-	d_7	-	-	b_7	0.000	0.12	d_7	0.011	1.39
b_8	-	-	d_8	-	-	b_8	0.000	-0.07	d_8	0.029	3.91
b_9	-	-	d_9	-	-	b_9	0.000	-1.67	d_9	0.018	2.47
obs.	4,564		obs.	4,564		obs.	19,345		obs.	19,345	
R^2	0.009		R^2	0.159		R^2	0.006		R^2	0.167	

Notes: The table presents the results (coefficients at the left hand side, z-statisic at the right hand side) of the vector autoregressive model described by equations (5.9) and (5.10), respectively. The estimates are based on all transactions for the 2008 Wimbledon final between *Roger Federer* (player 1) and *Rafael Nadal*. The interpretation of results is that inplay price changes are only affected by previous price changes and public knowledge innovations.

The results in Table 2 indicate that, on average, the implied relative winning probability of player 1 is increased by 0.0000378 immediately subsequent to a buy order. However, this effect is statistically not significant. It is instead only after another transaction that the midquote is raised by an additional 0.0000626 which is statistically significant on the 5% level. The positive signs on the coefficients for $x_t, \dots, x_{t-M}$ are in line with theoretical predictions; the more people buy player 1, the higher the relative winning probability for this player becomes.

A second important feature of our findings lies in the positive, and statistically significant effect from $x_t, \dots, x_{t-M}$ in the x_t Equation (3). This reflects the notion that buys tend to follow buys and that sells follow sells. Finally, it is

particularly interesting that we find negative signs for the coefficients on $dprob_{ij}^{t-1}, \ldots, dprob_{ij}^{t-M}$ in Equation (3). This could either be explained by inventory control effects (which would indicate that a sufficiently large number of individuals tend to place several bets before match begin); an individual with an inventory surplus for player 1 would lower the implied relative winning probability in order to elicit buys. As mentioned by Hasbrouck (1991a), an alternative explanation for this finding could be the price experimentation hypothesis suggested by Leach & Madhavan (1993) by which an individual sets quotes to optimally extract knowledge from other market participants.

Regarding the estimation of inplay transactions, we find again the positive autocorrelation in trades and a negative correlation between past price changes and trades. For the $dprob_{ij}^{t}$ series, however, the results look strikingly different from preplay transactions. The coefficients on $x_t, \ldots, x_{t-M}$ are not statistically significant. The interpretation of this finding is that inplay price changes are only affected by previous price changes and *public knowledge* innovations. This is perfectly in line with our expectations as all relevant information of the match can publicly be observed.

Our findings are also reflected in a graphical illustration of $\alpha_m(v_2^\tau)$ for preplay and inplay transactions of the 2008 final, displayed in Figure 1.

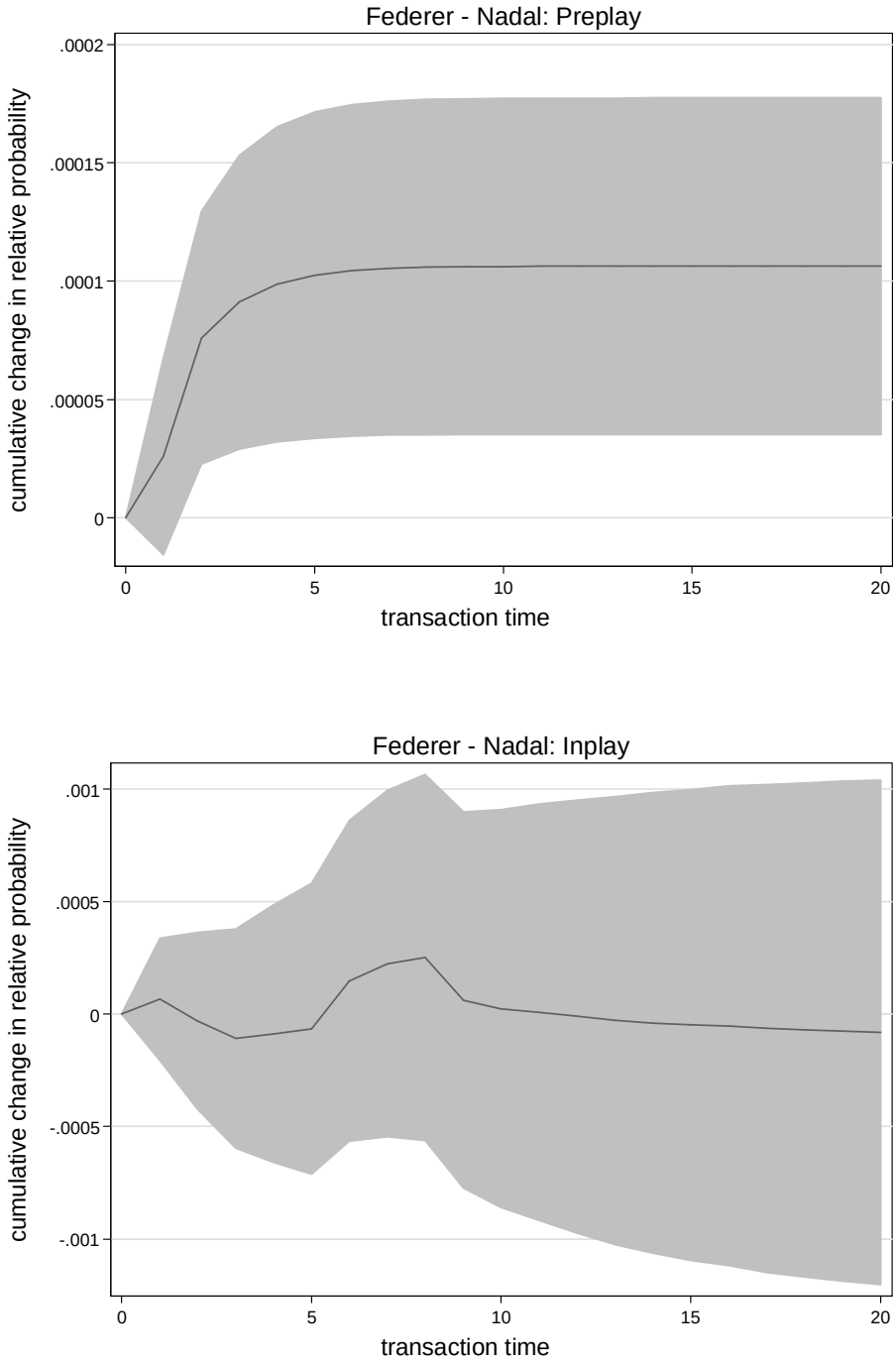


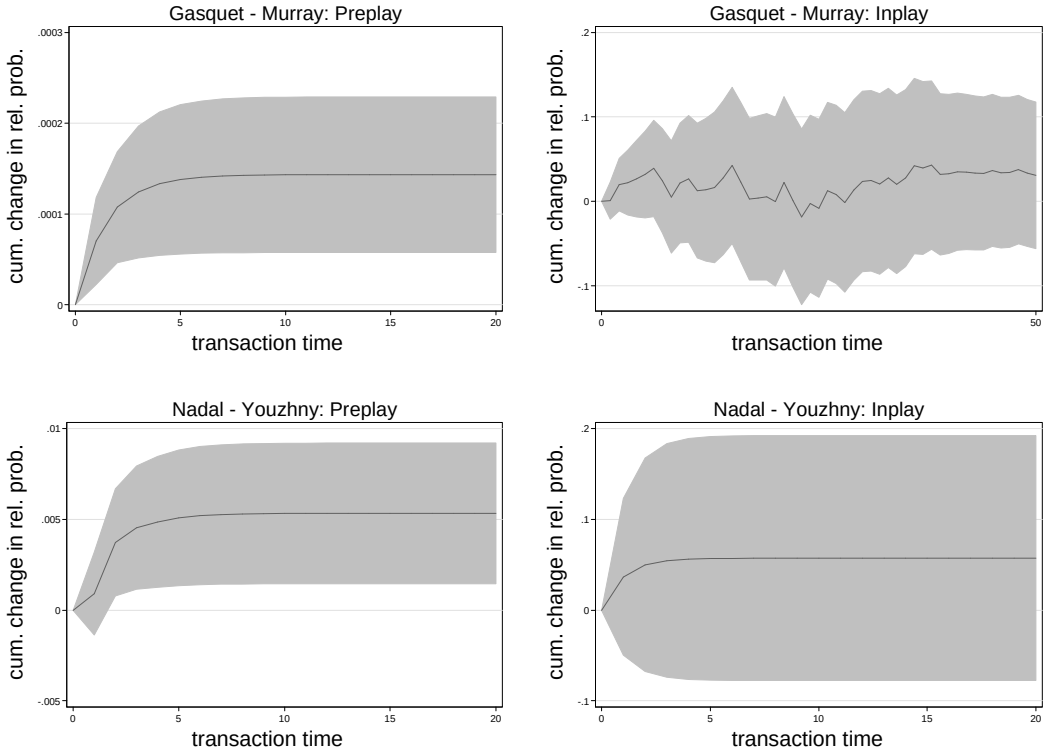
Figure 1: The probability revisions from signed trades for the 2008 Wimbledon final

The dynamic adjustment path implied by the bivariate vector autoregressive model described by Equations (9) and (10) is depicted for the 2008 Wimbledon final between *Federer* and *Nadal*. This is done for probability revisions caused by unexpected trades.

The Figure reveals the model's dynamic adjustment path which is inherently driven by the positive impact from trades on probability revisions, and the positive autocorrelation in the trade series. On top of Figure 1, the preplay series reveals that an unexpected, additional buy in $t = 0$ leads to a cumulative probability revision of roughly 0.0001 which is only fully incorporated after 4-5 steps, i.e., the adjustment is not instantaneous. In comparison, the 95% confidence interval for the inplay effect always includes zero.

To check for the relaibility of these results, we also estimated VAR models for five additional matches from varying playing rounds at the 2008 Wimbledon tournament. The associated graphical illustrations are displayed in Figure 2.

For the additional matches, the empirical results are less clear cut. For two of these matches, we find a long-term impact from private information innovations on price revisions in the preplay period. The impact from trades on quote revisions is not statistically significant for inplay transactions.



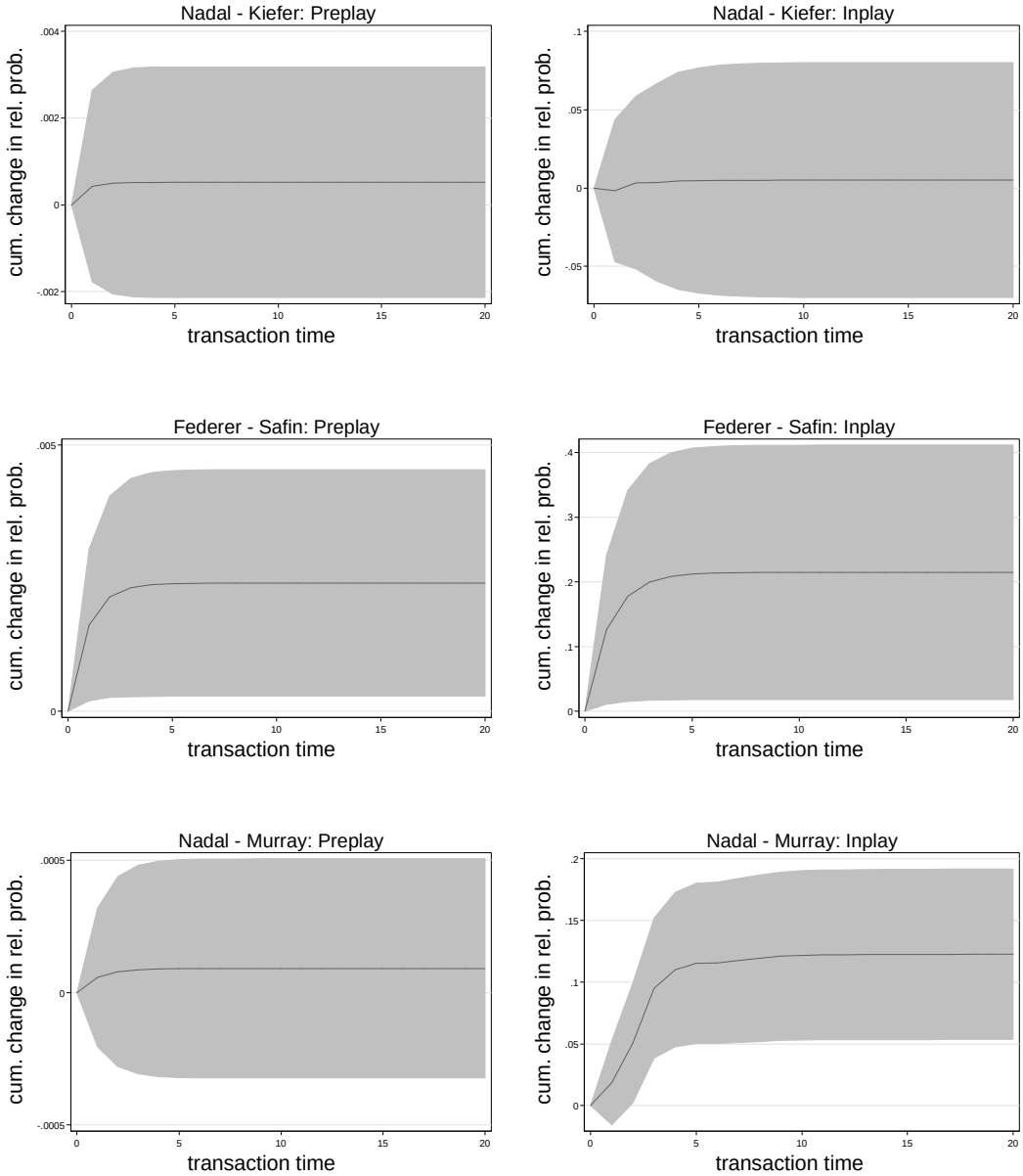


Figure 2: The probability revisions from signed trades for several matches

The dynamic adjustment path implied by the bivariate vector autoregressive model described by Equations (9) and (10) is depicted for a variety of 2008 Wimbledon tournament matches. This is done for probability revisions caused by unexpected trades.

One of the additional matches does not reveal a difference in the long-term impact of trades for preplay and inplay periods; the impact is statistically never different from zero (on a 5% level), but is also in line with our hypothesis. Finally, two matches appear to be somewhat special. In the quarter-final

match between *Nadal* and *Murray*, inplay price changes are correlated with historical trades, but preplay price changes are not. For the quarter-final match between *Federer* and *Safin*, there is a consistent correlation between trades and price changes, both, in preplay and inplay periods, although the effect is larger for inplay transactions. These results require further analysis and clarification. At this point, we only provide a very preliminary analysis of these effects; Table 3 presents an overview on the six matches analyzed so far.

Table 5.3
The economically most important matches of the Wimbledon 2008 tournament

match	round	rank (volume)	rank (transaction	preplay (volume)	preplay (transaction
Federer - Nadal	final	1 (49.1 mio)	1 (23,911)	6.3 mio	4,566
Federer - Safin	semi-final	6 (9.8 mio)	>10 (4,122)	3.9 mio	1,338
Nadal - Murray	quarter-final	3 (13.4 mio)	7 (7,482)	4.2 mio	3,178
Gasquet - Murray	fourth round	2 (24.6 mio)	2 (15,464)	1.2 mio	3,349
Nadal - Youzhny	fourth round	>10 (4.8 mio)	>10 (3,986)	1.2 mio	1,058
Nadal - Kiefer	third round	9 (7.0 mio)	>10 (4,975)	1.7 mio	1,703

Notes: The table presents the summary statistics on the most frequently traded matches of the Wimbledon 2008 tournament. Per match trading volumes (in GBP) and the number of transactions are given.

For each match, the total number of transactions and the total amount of matched volume are displayed. Based on this information, we ranked matches in terms of their economic value (volume matched) and market liquidity (transactions). However, a clear pattern cannot be detected from this Table; among the matches exhibiting the expected patterns are the final, two fourth round matches and a third round match. Although these matches include the top-2 matches with respect to volume and transactions, which may be consi-

dered somewhat illustrative, the other two matches were not among the top-10 in transactions.

Signed volumes

Within this section, we follow an extension by Hasbrouck (1991a) and incorporate additional trading characteristics into our model. An immediate extension arises from the concern that large trade innovations (measured by their volume) may influence subsequent price revisions differently than small trade innovations. Therefore, it seems appropriate to include signed trading volumes in the VAR estimation framework. As mentioned in section 5.4.1, the necessary information for this approach is readily available in the *Betfair* data. Based on the change of total volume matched for player i and j, we define the signed volume variable $V(x^t)$ as the net buying volume for player i[15]. A graphical illustration on $V(x_{ij}^t)$ for the 2008 final can be found in Figure 3.

[15] A numerical example may ease the understanding of the reader: Suppose between two transactions $t-1$ and t, bettors have backed 500 Pounds on player i, layed 1000 Pounds against player i and 1500 Pounds against player j. The net buying amount for player i is 1000 (=500 + 1500 -1000).

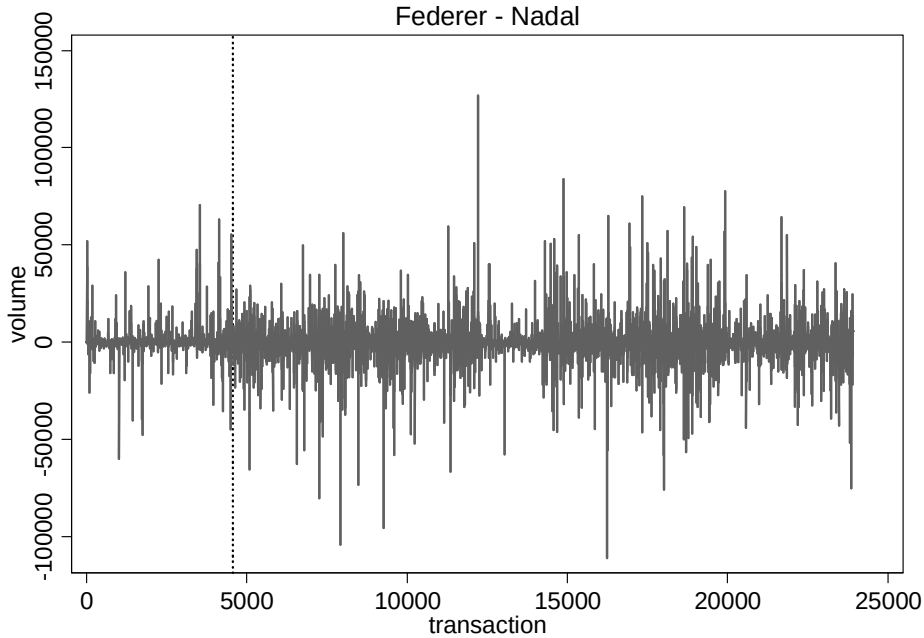


Figure 3: Signed volumes of the 2008 Wimbledon final

The signed volumes for the 2008 Wimbledon final are depicted for the preplay period (at the left hand side of the dotted vertical line) as well as during the match.

The corresponding estimation results are presented in Table 4. A graphical illustration for the implied $\alpha_m(v_2^\tau)$ of preplay and inplay transactions for the top-3 volume matches is displayed in Figure 4.

Table 5.4

Estimates of the bivariate vector autoregressive model for 2008 Wimbledon final (signed volumes)

preplay						inplay					
dprob			*x*			*dprob*			*x*		
a_1	-0.079	-5.34	c_1	-64.029	-9.00	a_1	0.039	5.39	c_1	-4.874	-15.77
a_2	-0.051	-3.40	c_2	-24.771	-3.45	a_2	0.021	2.86	c_2	-3.154	-10.12
a_3	-	-	c_3	-	-	a_3	0.048	6.55	c_3	-2.664	-8.52
b_1	0.000	1.22	d_1	0.359	23.60	b_1	0.000	0.47	d_1	0.253	34.53
b_2	0.000	2.03	d_2	0.084	5.70	b_2	0.000	-0.79	d_2	0.083	10.97
b_3	-	-	d_3	-	-	b_3	0.000	-0.41	d_3	0.055	7.21
obs.	4,552		obs.	4,552		obs.	17,665		obs.	17,665	
R^2	0.009		R^2	0.159		R^2	0.006		R^2	0.167	

Notes: The table presents the results (coefficients at the left hand side, z-statisic at the right hand side) of the vector autoregressive model described by equations (5.9) and (5.10), with the exeption that not only the trade direction but also the trade size is taken into account. The estimates are based on all transactions for the 2008 Wimbledon final between *Roger Federer* (player 1) and *Rafael Nadal*. The interpretation of results is that inplay price changes are only affected by previous price changes and public knowledge innovations

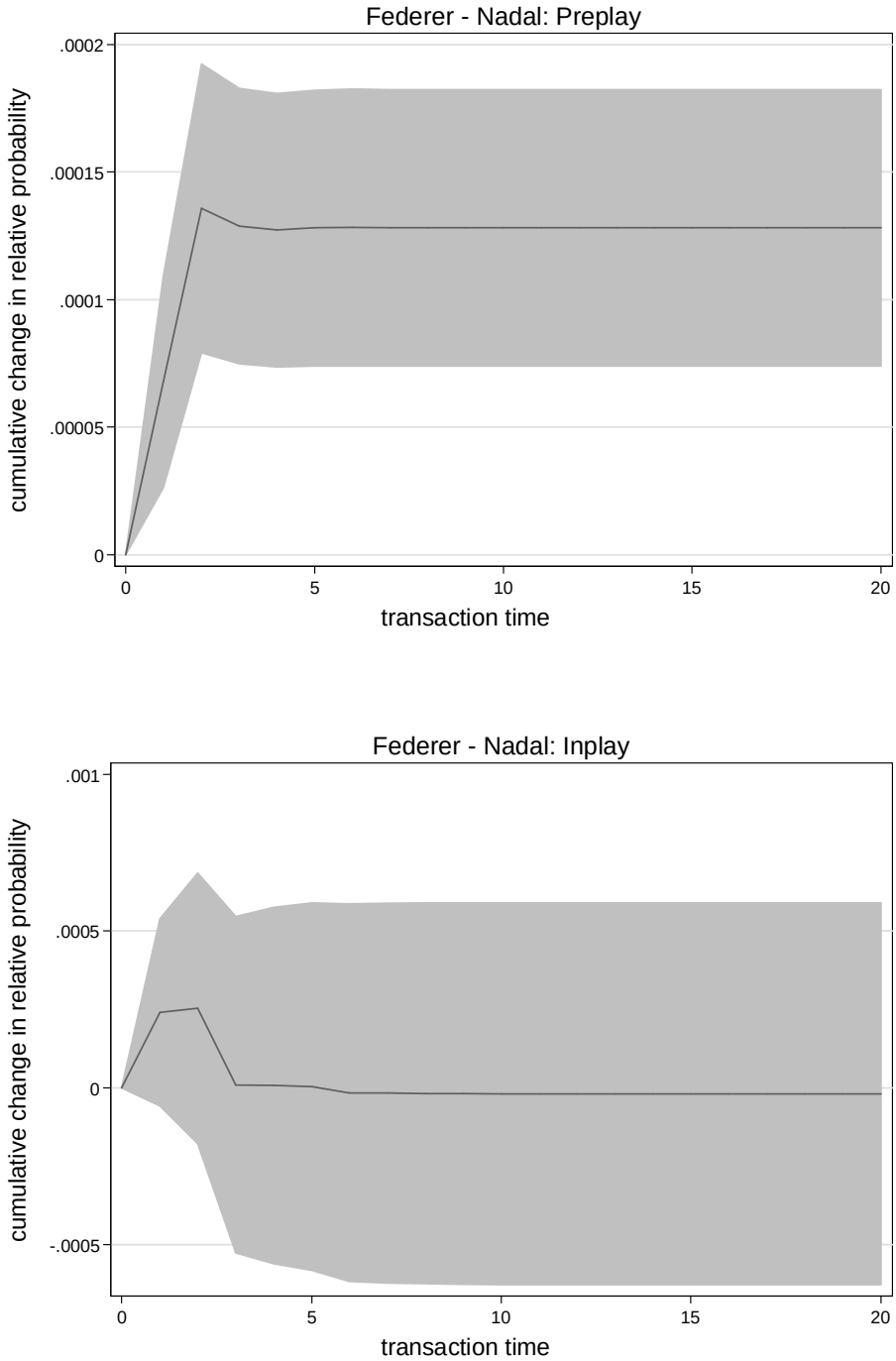


Figure 4: The probability revisions from signed volume for the 2008 Wimbledon final

The dynamic adjustment path implied by the bivariate vector autoregressive model described by Equations (9) and (10) is depicted for several 2008 Wimbledon tournament matches. This is done for probability revisions caused by unexpected signed volume.

Figure 4 shows that the findings on the 2008 final from the previous subsection remain valid; preplay price changes are correlated with previous transactions, but inplay price changes are not. Noteworthy, the preplay cumulative effect is now higher than than the (insignificant) average impact for inplay transactions. This supports our notion that trade size is relevant information for market participants, yielding a more realistic picture than signed trades.
Finally, Figure 5 shows that the application of signed volumes instead of signed trades does not improve the patterns of cumulative quote revisions for all matches.

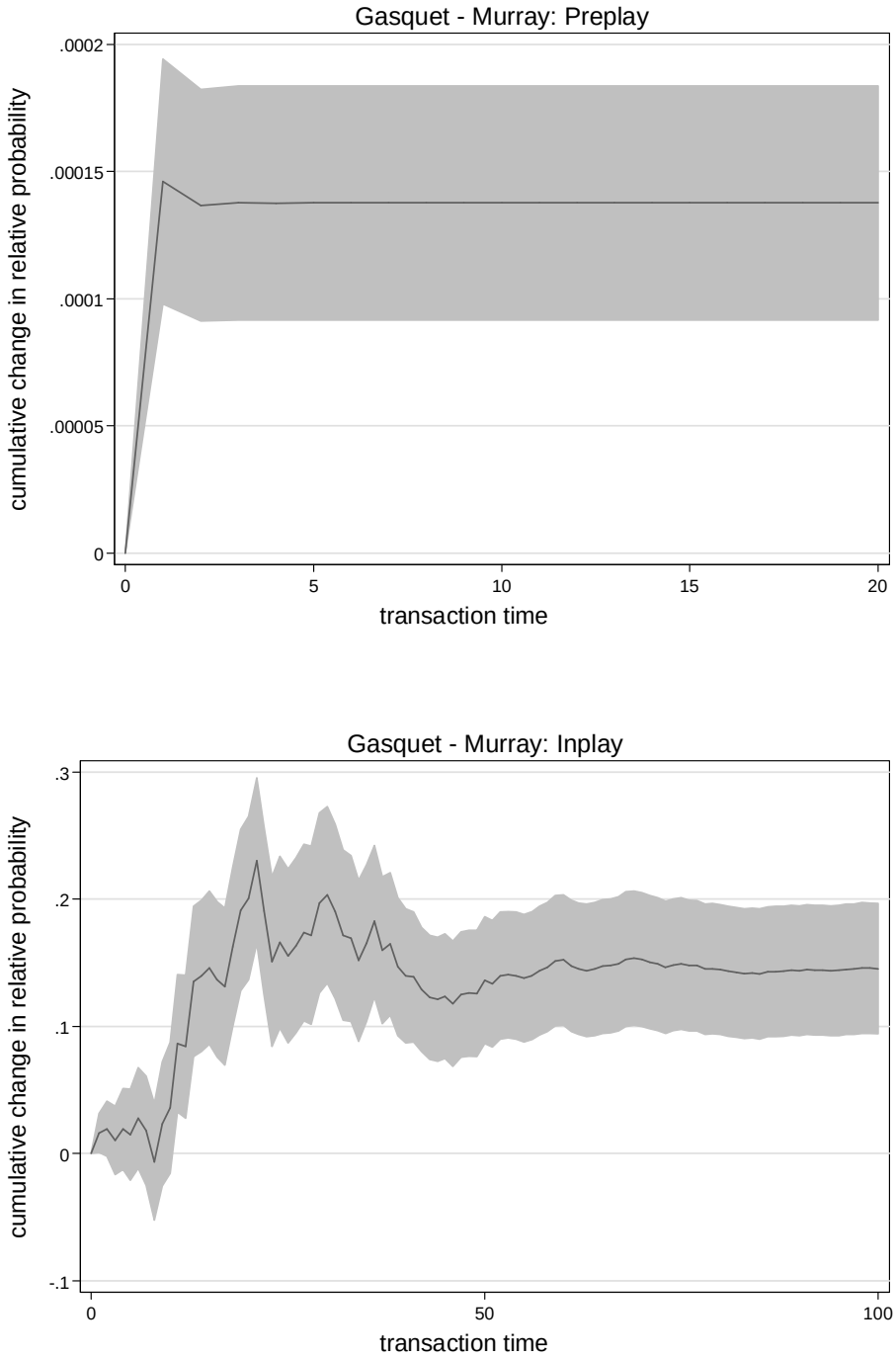


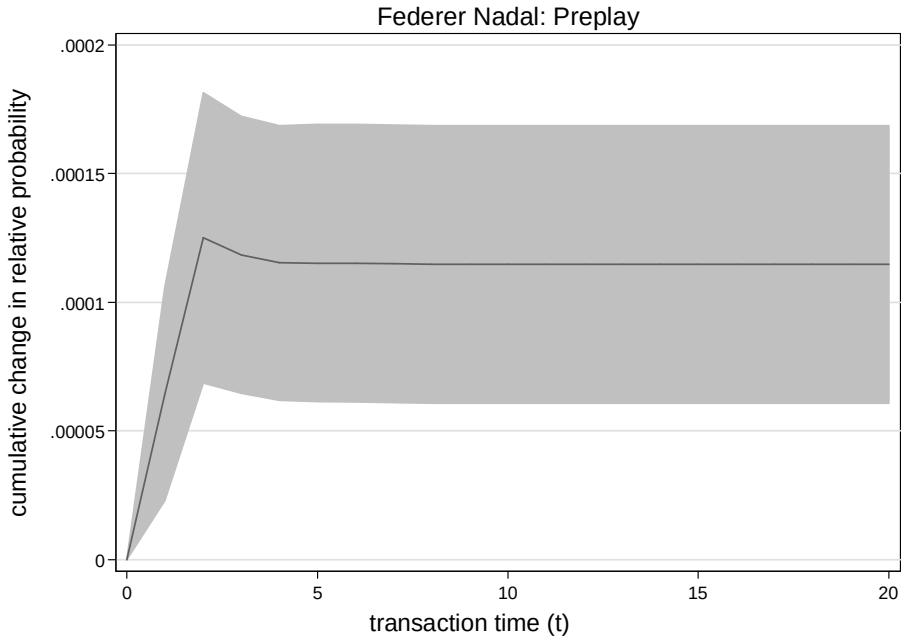
Figure 5: The probability revisions from signed volume for the *Gasquet-Murray* match

The dynamic adjustment path implied by the bivariate vector autoregressive model described by Equations (9) and (10) is depicted for several 2008 Wimbledon tournament matches. This is done for probability revisions caused by unexpected signed volumes.

For the *Gasquet - Murray* fourth round match, for example, we still find larger impacts from inplay transactions than from preplay transactions. In addition, for the *Nadal - Murray* match, a VAR model could not be fitted for inplay transactions as the SBIC suggested a lag order of 0. The reader should note, however, that the preplay transactions revealed the familiar, positive impact for cumulative quote revisions.

Nonlinearities in the response of quotes to trades

A final extension of our model includes the admission of nonlinearities with respect to the impact from trade size on quote revisions. To incorporate such nonlinearities into the VAR estimation approach, we follow the suggestion by Hasbrouck (1991a) and specify a model that is linear in nonlinear transformations of trade size. This leads to the formulation of a multivatiate VAR model for the variables set $\left\{dprob_{ij}^{t}, x_{ij}^{t}, V\left(x_{ij}^{t}\right), V\left(x_{ij}^{t}\right)^{2}\right\}$, where $dprob_{ij}^{t}$ is the probability change, x_{ij}^{t} is a trade indicator variable, $V\left(x_{ij}^{t}\right)$ is the signed trade volume and $V\left(x_{ij}^{t}\right)^{2} = x_{ij}^{t}\left|V\left(x_{ij}^{t}\right)^{2}\right|$. A graphical illustration of $\alpha_m(v_2^{\tau})$ for the 2008 final is presented in Figure 6.



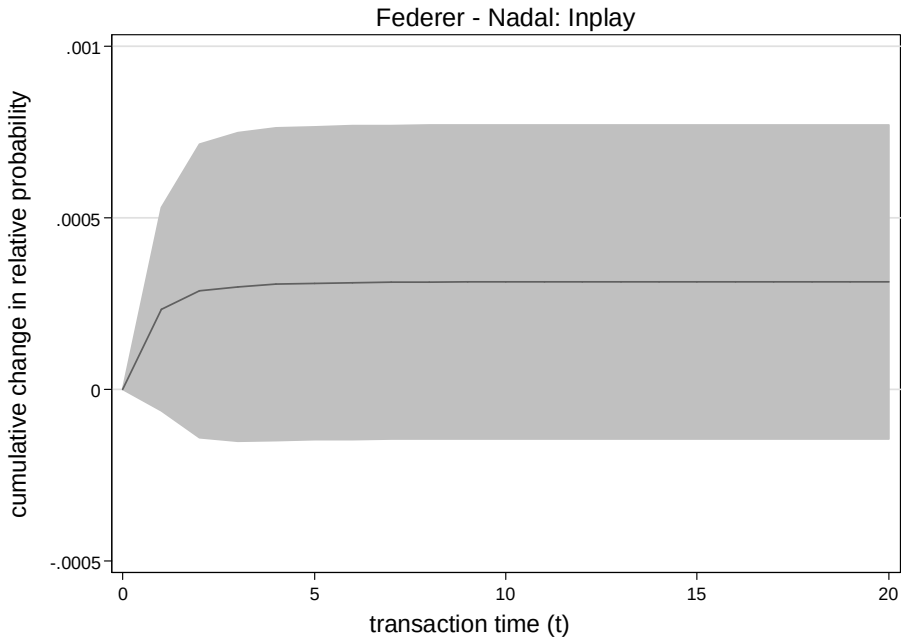


Figure 6: The probability revisions from the nonlinear transformations of trade size

The dynamic adjustment path implied by the bivariate vector autoregressive model described by Equations (9) and (10) is depicted for the 2008 Wimbledon final. This is done for probability revisions caused by the nonlinear transformations of trade size.

Interestingly, this specification leads to the same lag structure for preplay and inplay transactions ($M_1 = M_2 = 2$). As the expected pattern for $\alpha_m(v_2^\tau)$ can still be inferred from the estimation results, we are able to exclude the possibility that preplay and inplay differences in previous specifications are caused by varying lag structures.

Note, however, that nonlinearities should not have been expected for this type of betting market. Whereas large block trades may generally be „negotiated between participants in a loose ongoing business relationship“ (Hasbrouck, 1991a, p. 197) small trades may be handled relatively anonymous suggesting that the asymmetric information problem may be different for these two types of transactions. In the case of an online betting exchange, however, all trades have to be posted online, which means that - independent of a trade's size - the transaction procedure will always be the same. Under these conditions, it is not to be expected that the asymmetric information problem will be any different for large and small trades. This is what we find in our results (not given in this draft) for the 2008 final which we view as a simple plausibility check on the appropriateness of our estimation approach.

Conclusion

This paper contains a novel, direct testing procedure for the reliability of models on the information content of financial trades. In contrast to traditional studies, we apply a novel data set from the online betting exchange *Betfair*. This exchange has raised considerable interest from economists in recent years as it has become a major financial market (the number of daily transactions equals six times those at the London Stock Exchange). We argued that sports betting data - our analysis was based on full transaction records for selected professional tennis matches from the 2008 Wimbledon tournament - from this exchange are particularly suited to determine the information content of stock trades as trading occurs 'preplay', i.e., before match start, and 'inplay'. This allows us to fit separate models for times where the potential of informed investors tends to be lower (inplay) and those times where it tends to be higher (preplay). We then tested the hypothesis that measures on the information content of security trades exhibit larger values for preplay transactions than for inplay transactions. In case that we

cannot reject this hypothesis for a specific model, we would judge it to be reliable. Starting from the VAR estimation approach by Hasbrouck (1991a), which is based on price changes, signed trades, and signed volumes, subsequently, we find some evidence on the reliability of this model to measure the information content of security trades. For the 2008 Wimbledon Final, which exhibited the largest trading volume (USD 100 Mio), and was the most frequently traded in our sample, an additional preplay buy for player i in $t=0$ results in a permanent increase in his relative winning probability of 0.0001 (α =0.05). The impact for inplay buys, however, is not statistically significant. As all signs for the estimated coefficients in the VAR model for this match are in line with the documented findings in Hasbrouck (1991a), we are confident about the validity of our estimation approach. Altogether, the findings for four out of six matches clearly support our approach. For two matches, however, our approach leads to contradicting results.

Using information on signed transaction volumes instead of signed trades in a subsequent VAR model, we find supporting results, again, for the Final. A second match, however, shows a larger price effect from inplay transactions than from preplay transactions. Noteworthy, the matches supporting and contradicting our hypothesis are the same as in the signed trades framework.

To put our findings on more solid ground, however, the adoption of alternative models from the market microstructure literature is warranted to see whether different models come up with different, potentially contradicting results for identical matches. This is an important task for future research.

References

ADMATI, A. & PFLEIDERER, P. (1988): A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies*, *1*, 3-40.

BACK, A. (1992): Insider Trading in Continuous Time. *Review of Financial Studies*, *5*, 387-409.

BAGEHOT, W. (pseudonym) (1971): The Only Game in Town. *Financial Analysts Journal*, *8*, 31-53.

CHIANG, R. & VENKATESH, P. (1988): Insider Holdings and Perceptions of Information Asymmetry: A Note. *Journal of Finance, 43*, 1041-1048.

COHEN, L. & FRAZZINI, A. (2008): Economic Links and Predictable Returns. *Journal of Finance*, *63*, 1977-2011.

COPELAND, T. & GALAI, D. (1983): Information Effects on the Bid-Ask Spread. *Journal of Finance, 38*, 1457-1469.

CROXSON, K. & READE, J. (2008): Information and Efficiency: Goal Arrival in Soccer Betting. *Working paper*, Oxford University.

EASLEY, D. & O'HARA, M. (1987): Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics*, *19*, 69- 90.

EASLEY, D., ENGLE, R., O'HARA, M. & WU, L. (2008): Time-Varying Arrival Rates of Informed and Uninformed Trades. *Journal of Financial Econometrics*, *6*, 171-207.

EASLEY, D., HVIDKJAER, S. & O'HARA, M. (2002): Is Information Risk a Determinant of Asset Returns? *Journal of Finance*, *57*, 2185-2221.

EASLEY, D., KIEFER, N. & O'HARA, M. (1996): Cream-Skimming or Profit Sharing? The Curious Role of Purchased Order Flow. *Journal of Finance*, *51*, 811-833.

EASLEY, D., KIEFER, N. & O'HARA, M. (1997): One Day in the Life of a Very Common Stock. *Review of Financial Studies*, *10*, 805-835.

Easley, D., Kiefer, N., O'Hara, M. & Paperman, J. B. (1996): Liquidity, Information, and Infrequently Traded Stocks. *Journal of Finance*, *51*, 1405-1436.

Foster, D. & Viswanathan, S. (1990): A Theory of Interday Variations in Volume, Variance, and Trading Costs in Securities Markets. *Review of Financial Studies*, *3*, 593-624.

Foster, D. & Viswanathan, S. (1993): Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models, *Journal of Finance*, *48*, 187-211.

Foster, D. & Viswanathan, S. (1995): Trading Costs of Target Firms in Corporate Takeovers. *Advances in Financial Economics*, *1*, 37-57.

Glosten, L. & Harris, L. (1988): Estimating the Components of the Bid/Ask Spread. *Journal of Financial Economics*, *21*, 123-142.

Glosten, L. & Milgrom, P. (1985): Bid, Ask and Transaction Prices in a Spezialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, *14*, 71-100.

Glosten, L. (1987): Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices. *Journal of Finance*, *42*, 1293-1307.

Glosten, L. (1989): Insider Trading, Liquidity and the Role of the Monopolistic Specialist. *Journal of Business*, *62*, 211-235.

Hamilton, J. (1995): *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

Hasbrouck, J. (1988): Trades, Quotes, Inventories and Information. *Journal of Financial Economics, 22*, 229-252.

Hasbrouck, J. (1991a): Measuring the Information Content of Stock Trades. *Journal of Finance*, *46*, 179-207.

Hasbrouck, J. (1991b): The Summary Informativeness of Stock Trades: An Econometric Analysis. *Review of Financial Studies*, *4*, 571-595.

Hasbrouck, J. (2007): *Empirical Market Microstructure*. Oxford: Oxford University Press.

HUANG, R. & STOLL, H. (1997): The Components of the Bid-Ask Spread: A General Approach. *Review of Financial Studies*, *10*, 995-1034.

KYLE, A. (1985): Continuous Auctions and Insider Trading. *Econometrica*, *53*, 1315-1336.

LEACH, J. & MADHAVAN, A. (1993): Price Experimentation and Market Structure. *Review of Financial Studies*, *6*, 375-404.

LÜTKEPOHL, H. (2006): Vector Autoregressive Models. in: Mills, T. & Patterson K. (ed.): Palgrave Handbook of Econometics. New York: Palgrave-Macmillan.

MADHAVAN, A., RICHARDSON, M. & ROOMANS, M. (1997): Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks. *Review of Financial Studies*, *10*, 1035-1064.

MANSKI, C. (2004): Interpreting the Predictions of Prediction Markets. *Working paper*.

MCINISH, T. & WOOD, R. (1992): An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks. *Journal of Finance, 47*, 753-764.

O'HARA, M. (1995): *Market Microstructure Theory*, Oxford: Blackwell Publishers Ltd..

STOLL, H. (1989): Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests. *Journal of Finance, 44*, 115-134.