

Tick Test Accuracy in Foreign Exchange ECN Markets

Walid Ben Omrane*, Robert Welch

*Goodman School of Business
Department of Finance, Operation and Information Systems
Brock University*

January, 2013

Abstract

Recent changes in quoting activity have increased the allure of the Tick test because the quote rule and its variants require matching asynchronous trade and quote records. We find Tick test accuracy of forex USD/Euro trades in 2005/6 is 66%. The accuracy falls to 60% for zero tick trades (half the sample), with a disturbing asymmetry between buyer (61%) and seller (70%) initiated trades. We also observe extreme asymmetry for quote changes, where buyer accuracy is 96% (29%) for up (down) quote changes, respectively.

JEL classification: G10, F31.

Keywords: Tick rule; signed trades; foreign exchange ECN market; High-frequency data.

* Corresponding address: Goodman School of Business, Department of Finance, Operation and Information Systems, Brock University, 500 Glenridge Ave., St. Catharines, L2S 3A1, Ontario, Canada. Tel: 905 688 5550 x3920. Ben Omrane Email: wbenomrane@brocku.ca. Welch Email: bwelch@brocku.ca.

1. Introduction

Since most databases do not indicate the trade initiator, many research papers assign trade direction based on the tick test, the quote rule, LR (Lee and Ready (1991)), or EMO (Ellis, Michaely and O'Hara (2000)) with the latter two using a combination of the tick and quote rules. Consequently, the tick rule is used to sign some portion of the trade data in many articles which affects a significant portion of microstructure (order flow) research dependent on signing trades by these algorithms. An early example is Lyons (1995) who uses the tick rule to support an inventory control effect on price in the foreign exchange market for the USD/DM. The tick rule is simple to use as only trade data is required, which avoids the empirical errors caused by matching quotes to trades with delayed time stamps. This issue may be particularly important for newer data with its increase in quoting activity per trade. The tick rule can be fairly accurate under certain conditions but its accuracy declines for some markets (NASDAQ and electronic communication networks - ECNs) and for some subsamples such as zero tick trades. To our knowledge there has been no formal examination of the tick test in the foreign exchange market; consequently, this is our primary purpose.

Since our ECN trade/quote data on the USD/Euro exchange rate for 2005/6 is signed as buyer or seller initiated, we can definitively test the accuracy of the tick rule as well as examine the quote rule. However, all trades must occur on quotes and our results indirectly pertain to LR and EMO since they both use the quote rule for 'at quote trades'. From our sample of 679,162 forex trades, a subsample of 543,629 correctly match the previous best quote which implies that the quote rule is 80% accurate while the tick test is about 66% accurate. Although this subsample is evenly split between buyer/seller initiated trades, only 60.9% of buyer initiated trades are correctly classified by the tick rule; this is significantly less than the 70.3% for seller initiated trades. Half of this subsample are zero tick trades which only have a 61.2% accuracy rate. Unique to forex data are the 3.7% of trades that have a zero spread (bid equals ask), which can only be signed by the tick test, and have a 49% success rate.

Knowing the 'true' classification of our signed data not only implies more accurate tests of the signing algorithms but also more reliable empirical results for related research. This can be illustrated by the recent work of Chakrabarty et al (2012) that shows equity (NASDAQ) short sales are predominantly buyer initiated, which is contrary to conventional views (see Asquith et al (2010)). Chakrabarty uses the same 2 months in 2005 as Asquith, but has INET order data with a buy/sell indicator and designates the 'true' trade initiator as the last person to place their order chronologically. In general, not all transactions are classifiable by this method as Odders-White (2000) notes that 25.1% of her trade transactions were not classifiable because of missing order records. Another 'true' classification method requires knowledge of both traders' identities to assign direction. For example, a retail customer trading with a specialist, market

maker, or broker is assumed to be the initiator. However, not all trades can be signed with this method either; for instance, inter-broker or inter-dealer trades – see EMO who exclude 24.6% of their sample and Finucane (2000) who cannot identify trader types for 25% of TORQ trades. In addition, recent literature questions the usual assumption that intermediaries are passive liquidity providers – see Anand and Subramanyam (2008) and Chaboud et al (2012). A forex example of the importance of the true initiator is Bjonnes and Rime (2005), who do not find the inventory effect of Lyons (1995). They use the Odders-White classification (the latest chronological time stamp) instead of the tick rule to classify trade initiation. However, it should be noted that they have newer data from more brokers than Lyon’s data¹. Since most of the direct tests of these four signing algorithms rely on either the latest chronological order or knowledge of the participants to assign the ‘true’ classification, there is some error and loss of data. Consequently, knowledge of the true trade classification in our data is an important advantage for our study.

A related issue for the quote rule, LR, and EMO involves the matching of trade and quote data. Delayed matching is often necessary because (equity) trade data are not recorded as quickly as quotes. Depending on the type and age of the data, various delays from zero seconds (Bessembinder (2003)), one second (Piwowar (2006)), two seconds (Vergote (2005)), and five seconds (Lee and Ready (1991)) have been recommended to reduce signing errors when matching trades to quotes. EMO’s table 8 provides a breakdown of the asymmetric time delay between quotes and trades for the (1996/7) NYSE TAQ data. Examples of delays using the LR algorithm are Anand et al (2008) who use a one second delay in matching trades and quotes on 2002 TSX (ECN) data while Huh (2012) uses a two second delay for 1999 – 2009 NYSE data. Chakrabarty et al (2012) recommends returning to a one second delay from contemporaneous quotes recommended by Bessembinder (2003) and Petersen et al (2003) for newer data. Clearly, the tick rule avoids these potential matching errors. This matching issue may also affect the ‘true’ classification of buyer/seller initiated trades with proprietary data (see Boehmer et al (2007) and Bessembinder (2003)), including ECN order data using the ‘latest order submission time’ method of Odders-White. Holden and Jacobsen (2011) examine the difficulties in matching trades to quotes with modern equity tick data. Consequently, the tick rule has greater importance because it avoids these matching problems and needs to be scrutinized closely before using it to sign trades.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature; Section 3 discusses the signing algorithms and provides a brief summary of their tests and samples; Section 4 presents a brief description of the Hotspot forex data and some preliminary

¹ Berger et al (2008) also use the ‘latest time stamp’ to sign forex trades to relate order flow and exchange rates and find partial support for Evans and Lyons (2002 and 2008) with respect to high frequency rate movements.

results on spread size and trading frequency; Section 5 details the formal methodology of matching trades to quotes and cataloguing the seven main types of classification errors for the tick test; Section 6 discusses these tick rule misclassifications; Section 7 provides our main empirical results, including a detailed analysis of the most common errors and their role in the asymmetry between buyers and sellers; and Section 8 concludes the paper. The Appendix provides a brief summary of journal articles testing the signing algorithms.

2. Relevant Literature

Many studies use the tick rule because of its reported accuracy and simplicity of use. Lee and Ready (1991) state that when only trade data is available, the tick test performs remarkably well for 1988 NYSE equity data and that for mid quote trades, their simple model predicts 85% accuracy for the tick rule. However, if quotes are available they recommend a 5 second delay to match quotes with trades and use the LR algorithm which only uses the tick test for mid quote trades. They also conclude that for trades inside the quotes, the tick rule performs as well as the quote rule. This study was followed by Lee and Radhakrishna (2000) on 1991 TORQ data that found LR was 93% accurate on the 28.5% of trades in their sample that could be classified as buyer/seller initiated; this boosted the popularity of the LR rule over the tick test (see for example Chordia et al (2002), (2008), Anand et al (2008), Chung et al (2010), (2012) who all use LR). However, their discarded data is more likely to be misclassified (see Odders-White(2000) and Finucane(2000)), and the 93% is primarily due to 98% accuracy for trades at the quotes (76% of their sample) while mid quote trades (19% of their sample) use the tick rule (76% accurate) and the zero tick rule (60% accurate). Subsequent papers found reduced success for LR with a range of 72% to 85% (see appendix A). Since mid-quote trades have a lower success rate across these studies and LR uses the tick rule for these trades, some authors simply discard mid quote trades (such as Huh (2012)) and rely solely on the quote rule. However, omission of mid quote trades may also cause bias (see Finucane (2000)). Despite these results, LR is still quite popular.

The search for a more accurate signing rule is still an active research area. Chakrabarty et al (2007) find superior performance for a modified EMO rule for ECNs (INET and ARCAex) on NASDAQ stocks while Lu and Wei (2009) recommend a modified LR rule for the electronic order-matching Taiwan Stock Exchange. Both studies involve the tick rule for a portion of their data. For a stochastic analytical model of the tick rule see Perlin et al (2011), who determined its success rate based only on trades (a special case of their microstructure model is Roll (1984)) and then tested these predictions on Brazilian equities to find 72% accuracy for the tick rule.

Our 66% average accuracy for the tick test is lower than previous research in equity markets which are sensitive to the trading structure – specialist, market maker, or automated order matching. In particular, average tick rule accuracy for equity samples is almost 76% (the Appendix has a range of 72% to 83%). However, our 66% result is higher than the 59% for CBOE (Chicago Board of Options Exchange) option trades in Savickas and Wilson (2003), who note a large error rate for complex trades not executed on the automated execution system in their 1995 data. Poor subsample performance is common across many studies, and this could have a serious impact on research that depends on a reasonable level of accuracy for those trades. For example, correct classification rates are lower across all studies for zero tick trades (60% for NYSE equity in Odders-White (2000) and 77% in Finucane (2000)) and mid quote trades (63% Odders-White, 61% Ellis et al, and 66% Finucane). Furthermore, these average success rates may disguise other issues such as the asymmetry we found between buyer/seller initiated trades.

Indeed, because of the potential problems associated with the inclusion or exclusion of certain subsamples, many papers have found biases in microstructure research that depends on correctly signed trades. For example, Chakrabarty et al (2007) show the biases in estimating effective spreads and the price impact of trades for both LR and EMO. Petersen et al (2003) finds effective spreads overstate execution costs by 17% and recommend using EMO with contemporaneous quotes over LR. Finucane (2000) shows biases in signed volume and effective spreads (particularly if zero tick trades are excluded), but the tick test is not as biased as LR. Odders-White (2000) shows LR biases the transitory component of the spread as well as the reaction to bad earnings announcements. Ellis, Michaely and O’Hara (2000) also find that LR significantly overstates the effective spread, which is partly due to a poor (55%) classification accuracy for trades inside the spread which are prevalent for large trades, high volume periods, and ECN trades. Bessimbinder (2003) recommends EMO over LR because EMO has a smaller bias than LR in estimating trading costs, while Theissen (2001) finds LR and the tick rule create similar upward biases in the effective spread and its adverse selection component.

Petersen and Fialkowski (1994) note that accurate measures of transaction costs, such as the effective spread, are critical for tests of market efficiency, asset pricing models, and traditional market microstructure theories. Theissen (2001) also notes that since most data sets do not identify or provide information about the trader, much less the initiator, the accuracy of the trade classification algorithm is crucial to the validity of empirical research. Boehmer et al (2007) finds that LR inaccurate trade classification causes significant downward bias in estimating the PIN (probability of informed trades), yet Tay et al (2009) use LR to estimate PINs without correcting for estimation bias. Although most research papers recognize the existence of error rates for these trade classification algorithms, Tanggaard (2003) points out that no corrections or adjustments are attempted for the potential bias in the misclassification of the

buy/sell initiator. Some of the problems encountered with respect to errors relate to the type of security traded, poor sampling, lack of data, volatile markets, high volume periods, delayed recording procedures, and human interference in the trading process (including price improvements by specialists). In addition, not all trades may be classifiable as having an initiator, and in other cases the initiator may not always trade at a positive spread. To summarize, these biases manifest particularly for LR and by extension the quote rule which is LR's predominant component, but may also affect, to a lesser degree, the tick rule and therefore EMO which uses the tick rule for all trades not at a quote. It is clear that signed data from the source is the best alternative in order to avoid biases caused by signing rule errors.

How does the tick test perform relative to the alternatives – quote rule, LR, EMO? Since the tick rule is part of LR and EMO, only the quote rule is independent of the tick test; however, this comes with potential omission bias from the exclusion of mid quote trades. Recall that while the tick rule assigns a buy (sell) if the previous trade is lower (higher) respectively, the quote rule assigns a buy (sell) if a trade is closer to the ask (bid) respectively (mid quote trades are not signed). EMO uses the tick rule for all trades not 'at quote' while LR only uses the tick rule for mid quote trades. Consequently, EMO is more dependent on the tick rule than LR. The Appendix provides a summary of the average success rates across these algorithms where the most relevant comparison is between the tick and quote rules. It is clear that the quote rule has a higher success rate unless mid quote trades are counted as errors, in which case four papers find the tick rule to be superior. In fact, the papers by EMO, Theissen, and Chakrabarty find similar performances across all the algorithms; for instance Chakrabarty et al (2007) report much closer success rates for three algorithms on their 2005 INET ECN data – LR (74.4%), tick (75.4%), and EMO (75.8%) - and they recommend a modified EMO rule with a 76.5% success rate. Savickas and Wilson (2003) report the poorest relative performance for the tick rule² (59.4%) compared to EMO (76.5%), LR (80.1%), and the quote rule (82.9%) for 1995 CBOE options. Their tick test sample contained a higher percentage of index options which were poorly classified by all four algorithms

In general, ECN data has poorer classification success across all the algorithms and because of the dominance of electronic markets today, recent studies like Chakrabarty et al (2007) are perhaps more relevant. In fact, Jain (2005) reports 85 out of 120 exchanges are fully automated order matching systems as of 2001, and 101 of the leading exchanges have screen based electronic trading³. Therefore, unless there is a significant difference in performance between

² Except for Lu and Wei for the Taiwan Stock Exchange where their revised LR algorithm is the 'true' classification.

³ New York Times Business Day, January 1, 2011 by G., Bowley who notes that nearly all stock trading is done by high speed computers. Glosten (1994) models and discusses the advantages of electronic limit order books.

these signing methods, the tick rule is particularly appealing because it avoids the size, cost, and matching problems inherent in modern quote tick data.

The recent surge of computer trading in automated markets (ECNs) has substantially altered market microstructure (see Angel, Harris and Spatt (2010)), and necessitates a reexamination of old topics using newer data (see Holden et al (2011) regarding the breakdown of standard microstructure techniques). The recent large increase in quoting activity creates huge amounts of data requiring millisecond records. NASDAQ quotes per minute have increased by more than ten times between 2003 and 2009 (see Angel et al (2010)), while Hasbrouck and Saar (2009) find that 36% of nonmarketable limit orders are cancelled within 2 seconds. Hendershott et al (2011) find that algorithmic trading improves NYSE liquidity and quote quality. Stoll (2006) notes that ECNs are automatic, anonymous, low cost, and fast.⁴ Huang (2002) finds ECNs improve market quality by posting more informative quotes quicker than market makers, while Hendershott and Riordan (2009) find that algorithmic trading accounts for half the volume on DAX stocks. Since electronic markets are now central to modern day trading, ECNs are critically important to microstructure literature and hence, the trade initiator.

ECNs level the playing field and accommodate with more accuracy the huge increases in trading activity brought about by modern computer/telecommunication systems. The level playing field alludes to the changing role of market intermediaries (floor brokers/traders, specialists, and market makers) in traditional microstructure theory (see Kyle(1985) and Glosten (1987)) such as liquidity, bid-ask spreads, informed traders, adverse selection, and, more recently, their own informational advantage (see Anand et al (2008)). The recent increases in trading activity refers to the technological changes in the speed of computer assisted trading and the telecommunications advances in linking market participants to the current order book and trades - for example Boehmer (2005) and Garvey and Wu (2010) examine how speed impacts execution quality, while Chung et al (2012) study ECN's speed of convergence to market efficiency. In 2006, NYSE substantially reduced the execution time of market orders (from 10 seconds to less than one) via its Hybrid Market system in order to be competitive, deal with higher trading/quoting volume, and to qualify for the SEC Order Protection Rule for 'fast' markets (see Hendershott and Moulton (2011)). The Deutsche Boerse also reduced system latency from 50 ms to 10 ms in 2007 which increased liquidity and

⁴ Stoll also notes that small, inactive stocks may require intermediaries (such as market makers with affirmative obligations) to provide liquidity that is absent in limit orders. Erenburg and Lasser (2009) also note that limit orders (liquidity) can significantly decrease during times of high uncertainty typically associated with new information. In part this is due to a limit order's adverse selection risk inherent in its embedded option. Anand et al (2009) find that contracted liquidity providers improve market quality for low volume, wide spreads and high information asymmetry stocks.

efficiency (see Riordan and Storckenmaier (2012)). Therefore, a re-examination of signing rules for ECN data is necessary, especially for the tick test due to its simplicity.

In this regard, the papers of Aitken and Frino (1996) and Chakrabarty et al (2007) are quite relevant to this study. The former examines the tick rule on the automated order matching system for Australian equity (ASX) that has only 'at quote' trades, while the latter examines the tick, LR, EMO (plus their modified EMO) algorithms on newer NASDAQ ECN data. Our forex data, from Hotspot FX, is similar to the ASX as all trades occur 'at quotes', and it is also an ECN.

The work of Aitken and Frino (1996), hereafter AF, is of particular interest because their ASX data closely resembles key characteristics of our forex data. It is a fully computerized trading system which allows limit and market orders to be matched on price and time priority with no prominent role or privileges for market makers or brokers. Since all trades occur on quotes and the order record that initiated a trade always precedes it, there is very little true classification error by the quote rule. Their data also covers a 2 year period (ending in June, 1994) and they test the tick rule exclusively, finding a significantly lower success rate than predicted by Lee and Ready (1991). Although some of our results differ, there is a commonality worth pursuing in greater detail and we have structured our analysis along the same lines as AF. We have a lower overall tick test success rate of roughly 66% compared to AF's 74.4%; however our buyer initiated trades are poorly classified compared to seller initiated trades (60.9% versus 70.3% respectively), whereas AF only found significant asymmetry (with the seller poorly classified) when the mid quote has changed from the previous trade's mid quote. Half our sample are zero tick trades which only have a 61.2% accuracy rate; this is much less than the 68.5% success in AF whose sample has 75% zero tick trades. So, in spite of the older data, AF remain quite relevant since their data closely resembles the newer electronic data bases confronting researchers today.

Because the quote and tick rules are the building blocks in LR, EMO, and their variations, it is important to compare 'at quote' trades for these rules as one would expect the quote rule to be significantly better with a near perfect success rate. However, the tick rule does surprisingly well for some studies while the quote rule can have a rather high error rate. Specifically, the last column of the Appendix shows that the quote (tick) rule success rates for 'at quote' trades are about 89.4% (87.6%) in Finucane, 88.7% (83%) in Ellis et al, 80% (79%) in Chakrabarty, and 84% (63%) in Savikas and Wilson. Only Savikas and Wilson find a large advantage for the quote rule which persists for the retail automated electronic system (RAES). It is interesting that both rules improve their accuracy in RAES⁵, contradicting most research which has lower success rates for ECN type data. Some papers do not report success rates for 'at quote' trades, but we can obtain a sense for quote rule error in Odders-White who finds 89.6% success for 'at and outside the quotes' trades (the tick rule was about 75% accurate for all trades). Theissen

⁵ Complex index option trades not executed on RAES are part of the overall poor performance of the tick test.

reports 75.4% quote rule success for all trades except mid quote trades (the tick rule was about 72% accurate for all trades).

Contrast these results with automatic order matching systems that have nearly all their trades 'at quote'. AF's Australian data implies perfect accuracy for the quote rule which is their 'true' classification method. They find 74.4% accuracy for the tick test whose trades must be 'at quote' by definition. On the Taiwan Stock Exchange, Lu and Wei use their RLR (revised LR) rule as the true classification and find the 'at quote rule' is 92.68% correct with respect to RLR and over 94% of trades occur 'at quote'. They do not report tick rule accuracy separately for 'at quote trades', but find a tick rule accuracy of 74.2% for all trades. Perlin reports 72% tick test accuracy for Brazilian equity and all trades occur 'at quote' which implies perfect accuracy for the quote rule. Chakrabarty is the exception for ECN data (finding about 80% accuracy of 'at quote' trades for the quote rule) and it is noteworthy that only 58% of his trades occur on quotes. In this paper we know the true initiator and all forex trades occur on quotes; we find that 80% of trades correctly match the previous quote, and of the remaining 20% about 75% correctly match the post quote. This is roughly a 95% accuracy rate for the quote rule using both the pre and post quotes, and the tick test is about 66% accurate for the entire sample (including both pre and post quote matching trades). Without knowledge of the true trade initiator, we would not be able to reveal the quote rule's accuracy in the forex data. Consequently, for automatic order matching exchanges that have nearly all trades 'at quotes', the quote rule dominates the tick test. Clearly, the exchange trading system is an important consideration when choosing a signing algorithm.

Lastly, the results of several related papers are summarized according to the style of the exchange - specialist vs dealer - and the 'true classification' method which can all be found in the Appendix. The first exchange type is the specialist (auction market) that is typified by NYSE. Finucane (2000) and Odders-White (2000) both use 3 months of NYSE 1990/1 TORQ data (144 stocks) but different 'true' classification methods. The former identifies counter parties and assumes passive intermediaries to study 337,667 trades while the latter uses the latest time stamp to determine the trade initiator on 424,777 trades. Similar to Finucane, Theissen (2001) studies 15 German stocks for 21 days in 1996 using the Makler (specialist) as a passive trader to assign 'true' trade classifications. The second exchange type is the dealer (market maker) market typified by NASDAQ and the CBOE. EMO studied 313 NASDAQ stocks for the year ending in September, 1997 (2.4 million trades) using trader types and passive intermediaries to assign 'true' trade classifications while Savikas and Wilson (2003) used a similar 'true' classification system for equity and index options for the last six months of 1995 (2 million trades). All these studies lost a significant portion of their sample because 'true' classification of some trades was not possible due to missing quotes and/or time stamps, problematic matching of trades with quotes, and trader type ambiguity – reported percentages lost were about 25% by Odders-White, Finucane, and EMO, 30% by Savikas and Wilson, and 10% by Theissen. In contrast, the four studies with ECN data (see the Appendix) retain most of their data and use the latest time stamp to infer trade direction. This highlights a key feature of our ECN data in that the trade initiator is specified, reducing signing errors while keeping a large proportion of the trade data.

In summary, the tick test is tempting to use because it only requires trade data, which is much smaller in size than the ever increasing quote data found in modern 'tick' databases. The large number of electronically generated quotes requires multi-terabyte drives of highly compressed data that can be difficult to manipulate. The tick test also avoids the matching of asynchronous time stamps for trades and quotes as well as assumptions that certain types of traders are passive or that limit orders are always passive. However, in spite of these advantages our results indicate (at least for forex research) that matching trade data to quotes will successfully sign over 80% of the trades, while reliance on the tick test reduces this to 66% with an unexplained asymmetry between buyer/seller initiated trades. Hence, if signed trades or volumes are needed, then matching quote and trade data is necessary. Note that, in part, our results are due to the nature of the electronic (ECN) limit order book as trades can only occur at the bid/ask. This removes the ambiguity of trades inside (or outside) the quotes and, in particular, mid quote trades. This research has implications not only for market microstructure needing signed trades such as PIN (order imbalance), but also for market impact, event studies, and cross sectional returns (see Odders-White (2000) footnote 1). Consequently, the tick test, including its use in LR (for mid quote trades) and EMO (for trades not at the bid or ask), may have significant errors/biases that need to be recognized/corrected as argued in Tanggaard (2003). Notwithstanding our results, the NASDAQ (equity) ECN data of Chakrabarty (2007) - where only 57.5% of trades are at the bid or ask - shows that the tick test is as good as the quote rule and is a viable alternative for NASDAQ equity.

3. Signing Algorithms

The tick test (see Blume, Mackinlay and Tercker (1989)) requires only trade data and assigns 'buyer initiated' if the current trade price is higher than its predecessor (termed an uptick), while a downtick (current price is below its predecessor) is seller initiated. A zero tick occurs when the previous trade equals the current trade and the classification then depends on the trade before the previous trade. Therefore, a zero-uptick is a buy because the price change previous to the zero tick is an uptick (the trade prior to the previous trade was lower). A zero-downtick is a sell because the previous price change is a downtick. The reverse tick test uses the next (post) trade price to classify the current price (the current price is a sell if the next trade price is higher (uptick)). The tick rule and its reverse only differ when there is price continuation – two or more consecutive price changes in the same direction. For an example of this, see Figure 1 in Lee and Ready (1991) where a quote increase following a trade at the ask is correctly labelled as buyer initiated only by the tick rule. Hence, if quote revisions tend to be in the same direction as the previous trade (the quote increases after a buy), then the tick rule is expected to outperform its reverse.

The quote rule (see Hasbrouck (1988)) requires bid/ask quotes as well as trade prices and assigns a 'buy' to a trade which is closer to the ask than the bid (if the trade is closer to the bid, then it is a 'sell'). All trades midway between the bid and the ask (mid quote) are discarded. The 'at the quote rule' only assigns trades occurring at a bid or ask. The advantage of the quote rule over the tick test is that it incorporates additional information from the posted quotes, but at the cost of leaving mid quote trades unclassified. In addition, the differential time delay in recording trades vis a vis quotes creates errors in classification – even for proprietary data sets in determining the real initiator (see Bessimbinder (2003)). Consequently, the quote rule uses more information at greater complexity and cost but may not be significantly better than the tick test because of empirical issues.

LR (Lee and Ready (1991) recommend a combination of these approaches when quote data is available. They use the tick rule to classify mid quote trades and recommend the quote rule for all other trades. LR note that the tick rule works well for all trades inside the spread (not just mid quote trades), and should yield similar results to the quote rule for trades not at the ask or bid. In fact, it appears that the strength of the quote rule is its 98% accuracy (found by Lee and Radhakrishna (2000)) for trades at the quotes for the 28.5% of the TORQ data they could unambiguously sign. Their overall accuracy for the LR algorithm was 93%, but Odders-White (2000) notes that their eliminated data had a much lower success rate. Finucane (2000) also used TORQ data, which was augmented with several proprietary data sets⁶, and found an 83% accuracy rate for the tick rule and 84% for LR. Obviously the identification of the trade's true classification is essential for an accurate comparison of these algorithms as well as for the microstructure research dependent on their signed trades.

EMO (Ellis, Michaely and O'Hara(2000)) use the 'at the quote rule' for trades on quote (at the ask are buys while at the bid are sells) and the tick rule for all other trades. They find the quote rule correctly classifies 76% of the trades, the tick rule about 77%, LR about 81%, and their own EMO at about 82%. Their sample has 313 NASDAQ stocks over the year starting in September, 1996. Their proprietary data set has about 2.5 million trades and 627,000 quotes (BBO – best bid and offer), with a trader identity code and a buy/sell indicator. They use about 75% of the sample (1,834,000 trades) because they cannot sign trades between market makers, between brokers, or ECN trades. They assume that intermediaries (market makers and brokers) are liquidity providers to customers in order to sign trades. This assumption has been questioned in more recent work (see Anand et al (2008) and Chaboud et al (2012)) as it clearly affects the true classification of the initiator. EMO found that inside the spread trades were poorly classified;

⁶ Specifically the system order file (SOD) , the consolidated quote (CQ) and trade (CT) files and the audit (CD) file which all pertain to NYSE trades and quotes for Nov 1990 to Jan 1991. This data was used to determine the actual initiator but, even so, about 25% of the TORQ data could not be signed – mainly because one (or both) trader types were not identified.

however, the tick rule was better than the quote rule (61% versus 55%), and, consequently, they proposed their EMO rule as a better alternative to LR.

Chakrabarty et al (2007) modify EMO by using the tick rule for trades above the ask, below the bid, and close to the mid quote (plus or minus 20% of the bid-ask spread). In other words, they use the quote rule for trades at or below the ask by 30% of the bid-ask spread (designated a buy) and denote a sell for trades at or above the bid by 30% of the bid-ask spread. They use NASDAQ ECN data (2 months in 2005) with ITCH data that has a buy/sell indicator and then match trade and quote data, using the latest time stamp to determine the initiator. The algorithm success rates are 74.4% (LR), 75.4% (Tick), 75.8% (EMO), and 76.5% for their modified EMO rule. Interestingly, they only find 58% of their ECN trades occur at the quotes, which may explain these generally lower success rates compared to previous research (Chakrabarty success rates for 'at quote' trades are about 79% for the tick test and 81% for the quote rule). However, they note a significant number of trades outside the NBBO quotes (national best bid and offer); this was also observed by Stoll and Schenzler (2006) for NASDAQ stocks. This could result from the lack of a 'trade through' rule at the time of their study or the matching process used for the creation of the NBBO series. While Odders-White could not sign 25% of her 1991 NYSE trade data due to the absence of order execution records (no date and time to match to the trade), this did not seem to be a problem for the 2005 ECN data on NASDAQ stocks. In any event, the knowledge of the true initiator in our forex data is important because it removes some of the uncertainty in regards to evaluating the tick test and the quote rule.

Lu and Wei (2009) propose a modified LR rule for the Taiwan Stock Exchange (TSE) which is similar to the order matching system of the Australian Stock Exchange (ASX) that is based on price and time priority. They adjust for missing bid or ask quotes due to low liquidity or trading price limits. Since 94% of trades occur at the quote, they recommend the quote rule (modified for one sided quotes) followed by the tick rule for all other trades which they refer to as the revised LR rule (RLR). They use contemporaneous matching (a zero second delay) of the trade file to the order file. Since the trade initiator is not known, they compare the other trading rules to RLR and found that while the tick rule classifies the most trades, its success rate (74.2%) was worse than the quote rule (93%), LR (96.5%), and EMO (95%) with respect to the 17 million RLR signed trades in 2006.

Perlin et al (2011) use a Markov chain process to model trades and predict the correct percentage of tick rule classifications. One implication of their work is that the implied bias of the tick test can be evaluated even though the researcher only has trade data. This could address the consequences of trade sign misclassification (Tanggard (2003) and Boehmer et al (2007)). They also used trade data on 15 of the most liquid Brazilian stocks in 2009 to find that the tick test is about 72% accurate, which closely matched their model predictions. The Brazilian stock exchange is an electronic limit order book structure with price and time priority where every trade has a matching buy and sell order.

4. Hotspot Data

The Hotspot forex system started trading in 2002 and was the first multibank platform to allow client quotes with a transparent limit order book for the dealer to client spot forex market. Clients are endowment and pension fund managers, treasury departments, hedge funds, CTAs, trading firms, banks, and other retail traders who can all trade directly with any of the member banks or other clients. Clients and institutions can submit market and limit orders which are displayed anonymously in a central order book. A distinguishing feature of this system compared to the traditional 'request for quote' (RFQ) model is the client to client trades which occur with full knowledge of the order book from all quoting parties. In the RFQ model, clients only deal with and see the quotes from their dealer's bank.

In Hotspot, all trades are cleared through a dozen clearing banks and prime brokers. Counterparty information is only given to the client's prime broker/clearing bank. Client to client trades account for substantial volume and member banks do not trade directly with each other. In essence, the member (market maker) banks provide executable prices and credit while clients also provide liquidity through their orders. Both banks and clients can actively trade (market orders) and all trades/quotes are treated in the same anonymous fashion. All clients/brokers see the order book - bid/offer quotes and associated depth - with no special advantages to any traders (equal information access)⁷. Hotspot is a 'commission only' business model (usually a fraction of a basis point (bp) is charged to both counter-parties) with no monthly subscriber charges, technology, licensing, or other fees. Typically, most trades are executed within 10 milliseconds. Client to client trades can be nearly half the daily volume with a system enforced 'best price execution' and real time audit trail. This allows for the lowest trading costs (best prices) with full information. Clients that do not qualify for a direct PFI (participating financial institution) line access Hotspot via a general account through a forex broker/PFI⁸. According to a 2006 Euromoney FX poll, Hotspot was second in total market share for multibank platforms and first in execution speed for spot forex. It important to note that by 2006, forex ECNs were rapidly gaining market share from the traditional dealer market RFQ model.

We have the quotes and trades for the USD/EURO from Hotspot in 2005 and 2006. The following sample trade record matches its previous quote – the second quote is the post quote.

⁷ Bank client credit exposure is only to its clearing corporation – never to another client or bank.

⁸ Participating financial institutions include Bank of America, BNP Paribas, Citibank, Deutsche Bank, Goldman Sachs, HSBC, JP Morgan, Morgan Stanley, Newedge, Prudential Financial, Societe General, UBS, and other clearing forex banks.

<i>Trade Date</i>	<i>Start Time</i>	<i>Currency</i>	<i>Initiator</i>	<i>Rate</i>
20050103	00:47:32:270	USD/EUR	buyer	1.34430

<i>Quote Date</i>	<i>Quote Time</i>	<i>Quote Bid</i>	<i>Bid Size</i>	<i>Quote Ask</i>	<i>Quote Size</i>
20050103	00:47:32.060	1.34420	5,000,000	1.34430	5,000,000
20050103	00:47:32.388	1.34420	5,000,000	1.34430	1,000,000

A buyer initiated trade occurred 47 minutes, 32 seconds, and 270 thousandths of a second after midnight on January 3, 2005. It matched its previous quote at the ask at \$1.3443/EURO which was posted 210 thousandths of a second before the trade. The post quote occurred 118 thousandth of a second later, and the trade size was 4 million Euros which is the difference between the pre and post ask quote sizes. In general, we are not able to determine the trade size with certainty for all trades. Minimum quote size on Hotspot is 1 million euros.

We first match each trade's price with the best previous quote's bid or ask, and if the buy/sell initiator indicates agreement with the quote - a buyer (seller) initiator price equals the ask (bid) - then the trade is included in our main sample. This occurs for 543,629 trades which is approximately 80% of the total sample of 679,162 trades. For the 20% which had no agreement (a 'mismatch' of 135,533 trades), we then compare the trade price with the nearest post quote. About 75% of these trade prices correctly agree with the post quote's bid/ask, and as a result, about 95% of the trades correctly match their pre or post quote. The remaining 5% all have the conflicting information that the buyer/seller initiated trade price is the same as the post quote's bid/ask respectively. To be conservative, we restrict our sample results to the 80% of trades that correctly match the previous quote.

Some descriptive statistics for spread size and time between trades are presented in Tables 1 and 2 respectively for the main sample which matched the previous quote (543,629 trades). Spread size in Table 1 has a mean of 1.09 bp (1 basis point = .0001\$) and a standard deviation of 0.9 bp with the bulk of the observations being 0.5 bp (23.3%) and 1 bp (46.8%). Note that half bp pricing started in August of 2005. There are zero spreads for which the bid equals the ask (3.7% of the sample) and these have the worst success rates for the tick test – 43.14% (buyer) and 52.16% (seller). The asymmetry between buyer and seller is strongest for zero spreads, but weakens as the spread size increases and in fact reverses for spreads larger than 2 bp (only 2.55% of the sample). Notice that the success rates for both buyer and seller increase monotonically to 77.36% (buyer) and 75.67% (seller) as the spread size increases. With the average spread size close to one bp, there is not much room for spreads to decrease in size.

[Insert Table 1 about here]

The amount of time between trades (duration) for the main sample in Table 2 has a mean of 18.63 seconds and a standard deviation of 40.67 seconds. About 45.5% of the trades occur within 3 seconds of each other, while 41.56% have a duration greater than 10 seconds. Correct classification by the tick test is less than 3% for trades occurring within 0.05 seconds of each other (only 0.16% of the sample). This improves dramatically to 75.95% (seller) and 49.77% (buyer) for durations between 0.05 and 0.1 seconds (8.99% of the sample). Then, seller side accuracy declines with longer durations while the buyer side hovers between 59% and 64%. Again we notice significant asymmetry between buyer and seller initiated trades for short durations under one second (35.68% of the sample), but that weakens for durations between 1 and 60 seconds and then reverses for durations exceeding 10 minutes. The largest difference between buyer and seller percentage success is 26.18% (75.95% - 49.77%) for trades within 0.05 and 0.1 seconds of each other, and the asymmetry is still 9.23% for trades within 2 to 3 seconds. The average time between a trade and its previous quote is 1.13 seconds with a standard deviation of 46.45 seconds.

[Insert Table 2 about here]

In summary, Tables 1 and 2 reveal that most of the main sample and its asymmetry between buyer/seller occurs for small spreads (within 1 bp – about 73.8% of the sample) and during periods of fast trading (less than 3 seconds between trades – about 45.5% of the sample). However, it is not clear why buyer initiated trades are more difficult to classify and we examine this further in Section 7.

5. Methodology

We match a trade to its most recent previous best quote. Let t_q and t_d be the quote and transaction times respectively. Q is the quote price and Q^M is the matched quote at time t^* with the corresponding deal (trade) D , where $Q^M = (Q_{t^*,Bid}^M, Q_{t^*,Ask}^M)$,

$D = (D_{t_d,Buy}, D_{t_d,Sell})$ and t^* is the matching time, such that;

$$Q_{t^*}^M = \{Q_{t^*} | [t^* = \text{Min}(t_d - t_q); t_q \leq t_d]\}. \quad (1)$$

We consider signed trades where both buyer and seller initiated transactions are known. To check for the performance of the quote rule, we need to make sure that the buyer initiated trade matches the previous ask quote while the seller initiated trade corresponds to the bid quote. We can then compute the indicator QR , which is the percentage accuracy of the quote rule in classifying the trades:

$$QR = \frac{I' \cdot \mathbb{1}_{\left\{ \frac{Q_{t_d}^M, Ask=1}{D_{t_d, Buy}} \right\}} + I' \cdot \mathbb{1}_{\left\{ \frac{Q_{t_d}^M, Bid=1}{D_{t_d, Sell}} \right\}}}{I' \cdot \left(\mathbb{1}_{\{D_{t_d, Buy}\}} + \mathbb{1}_{\{D_{t_d, Sell}\}} \right)} \quad (2)$$

Where I is a vector of ones and $\mathbb{1}$ is an indicative vector that takes the value of one if the condition specified above is satisfied, otherwise it is zero.

In order to check the accuracy of the tick rule, we consider the trade vector D and we classify trades as a buy (B) or sell (S) with respect to the trade price movement:

$$\begin{cases} Trade_{t_d}^B = \mathbb{1}_{\{D_{t_d} > D_{t_{d-1}}\}} + \mathbb{1}_{\{D_{t_d} = D_{t_{d-1}}; D_{t_{d-1}} > D_{t_{d-2}}\}} \\ Trade_{t_d}^S = \mathbb{1}_{\{D_{t_d} < D_{t_{d-1}}\}} + \mathbb{1}_{\{D_{t_d} = D_{t_{d-1}}; D_{t_{d-1}} < D_{t_{d-2}}\}} \end{cases} \quad (3)$$

Where $Trade_{t_d}^B$ and $Trade_{t_d}^S$ are the trades classified as buy or sell respectively according to the tick rule.

The tick rule classification is successful if $Trade_{t_d}^B = D_{t_d, Buy}$ and/or $Trade_{t_d}^S = D_{t_d, Sell}$ otherwise it fails because of misclassification errors.

We identify at least six misclassification errors that occur when using the tick rule to assign trade classifications (see Figure 1 and its explanation in section 6). We detail these buyer initiated errors for downticks (E_1 and E_2) and zero downticks (E_3 to E_6) and note that there are six symmetrical errors for upticks (not presented). A similar set of 12 scenarios exists for seller initiated trades.

E_1 is the classification error type 1 that occurs on a downtick for consecutive buy trades when the ask is decreasing:

$$E_1 = \{Q_{t^*, Ask}^M < Q_{t^*-1, Ask}^M \mid Q_{t^*, Ask}^M = D_{t_d, Buy}; Q_{t^*-1, Ask}^M = D_{t_{d-1}, Buy}\} \quad (4)$$

E_2 is the classification error type 2 that occurs when a sell is followed by a buy and the ask falls below the previous bid :

$$E_2 = \{Q_{t^*, Ask}^M < Q_{t^*-1, Bid}^M \mid Q_{t^*, Ask}^M = D_{t_d, Buy}; Q_{t^*-1, Bid}^M = D_{t_{d-1}, Sell}\} \quad (5)$$

The classification error type 3, E_3 , occurs on a zero downtick for consecutive sell trades when the last (t-1) bid falls below the previous (t-2) bid, followed by a current (t) buy when the current ask falls to the bid on the previous (t-1) trade:

$$E_3 = \{Q_{t^*,Ask}^M = Q_{t^*-1,Bid}^M < Q_{t^*-2,Bid}^M | Q_{t^*,Ask}^M = D_{t_d,Buy}; Q_{t^*-1,Bid}^M = D_{t_d-1,Sell}; Q_{t^*-2,Bid}^M = D_{t_d-2,Sell}\} \quad (6)$$

Also occurring on a zero downtick is classification error type 4, E_4 . This error occurs when a buy trade (t-2) is followed by a sell (t-1) and the last bid (t-1) is less than the previous ask (t-2), is followed by a buy (t) when the current (t) ask falls to the previous trade's bid (t-1):

$$E_4 = \{Q_{t^*,Ask}^M = Q_{t^*-1,Bid}^M < Q_{t^*-2,Ask}^M | Q_{t^*,Ask}^M = D_{t_d,Buy}; Q_{t^*-1,Bid}^M = D_{t_d-1,Sell}; Q_{t^*-2,Ask}^M = D_{t_d-2,Buy}\} \quad (7)$$

In the case of consecutive buys when the last (t-1) ask quote falls below the previous (t-2) ask, is followed by a buy (t) when the current (t) ask falls to the previous (t-1) ask, error type 5, E_5 , is generated:

$$E_5 = \{Q_{t^*,Ask}^M = Q_{t^*-1,Ask}^M < Q_{t^*-2,Ask}^M | Q_{t^*,Ask}^M = D_{t_d,Buy}; Q_{t^*-1,Ask}^M = D_{t_d-1,Buy}; Q_{t^*-2,Ask}^M = D_{t_d-2,Buy}\} \quad (8)$$

Finally, E_6 occurs on a zero downtick if a sell (t-2) is followed by a buy (t-1) when the last ask (t-1) falls below the previous (t-2) bid, is followed by a current (t) buy when the current (t) ask falls on the previous (t-1)ask:

$$E_6 = \{Q_{t^*,Ask}^M = Q_{t^*-1,Ask}^M < Q_{t^*-2,Bid}^M | Q_{t^*,Ask}^M = D_{t_d,Buy}; Q_{t^*-1,Ask}^M = D_{t_d-1,Buy}; Q_{t^*-2,Bid}^M = D_{t_d-2,Sell}\} \quad (9)$$

6. Tick Rule Misclassifications

Following AF, we graphically present the six types of tick test buy errors from the previous section in the panels of Figure 1 when there is a price decrease (errors for a price increase are symmetrical and are not presented). Panels A and B are downticks, while the rest (panels C to F) are zero downticks. Panel A (error 1) considers the current (time t) buy trade at the ask on the far right of the time axis and six possible previous buy trades to its left at time $(t-1)$. The bid, ask, and the mid quote are displayed for each possibility, including the last one where the bid = ask (a zero spread). Since the previous trade $(t-1)$ is higher than the current trade (t) , the tick rule denotes a sell. This is an error since the current trade (t) is a buy at the ask and we denote this as error type 1 to coincide with AF as a buy - buy down tick error.

Panel B (error 2) is a downtick sell-buy error. The current trade is a buy at the ask and the previous trade is a sell at the bid (the other previous trade is a zero spread trade with bid = ask). Since the current trade price is less than the previous price, the tick rule assigns a sell initiator – this is error type 2 in AF, a sell-buy downtick error. If the previous trade was a buy (at the ask) this would be the type 1 error of panel A as the downtick would cause the tick test to designate a sell. The difference between panels A and B is the larger decrease of the current quote. Specifically, the current ask is less than the previous bid (panel B), while it is only lower than the previous ask in panel A.

[Insert Figure 1 about here]

Panel C (error 3) is a zero downtick error that AF call type 3 error (sell –sell –buy). The current (t) trade is a buy equaling the previous trade at $(t-1)$ which is a sell, creating a zero tick, so going back to $(t-2)$ there is another sell greater than the sell at $(t-1)$ and this is the downtick. Hence, the tick test assigns a sell to the current trade which is an error since the current quote has fallen relative to the previous quote; it is actually a buy. Panel D (error 4) is similar except that the trade at $(t-2)$ is a buy exceeding the sell at $(t-1)$, creating the down tick prior to the zero tick between t and $(t-1)$. Again, the tick rule assigns a sell due to the previous down tick –a type 4 error in AF (buy-sell-buy). In both panels C and D the current quote has fallen with a buy at the ask, creating the zero tick.

Panel E is AF's type 5 error (buy-buy-buy). The current buy trade at t is at the same ask as the previous buy at $(t-1)$, creating the zero tick but the trade at $(t-2)$ is another buy at a higher ask (the ask price fell between $(t-2)$ and $(t-1)$) creating a downtick. Therefore, the tick rule assigns a sell at t which is an error. Panel F is AF's type 6 error (sell-buy-buy). The only difference to panel E is a sell instead of a buy at $(t-2)$ which is still higher than the buy at $(t-1)$ because the $(t-1)$

quote has fallen more than it did in panel E, creating a downtick. Consequently, the tick rule assigns a sell which is an error.

AF note that as quotes decrease in a falling market, the tick rule will bias initiator classification in the direction of the market movement. This is evident in Figure 1 as all the errors occur in a falling market which causes the tick test to classify a sell on the down or zero downtick (an error as the current trade occurred at the ask which is a buy). Similarly, in a rising market the tick test will bias towards a buy classification as quotes move upward which creates an error if the current (t) quote has increased enough to be a sell at the new bid. Thus, for quote changes from the previous trade's quote, AF expect to see poorer misclassification by the tick test – which they did in fact observe. Interestingly, our forex results are not as neatly explained, and we discuss this in greater detail in the next section.

7. Empirical Results

The Hotspot data for 2005/6 has 679,162 trades and includes the buy/sell trade initiator. About 80% of these trades (543,629) matched correctly with the best current quote. Specifically, a trade with a buy (sell) initiator had the same price as the ask (bid) of the previous (best) quote. It is worth noting that for the 20% of trades (135,533) which do not match the previous quote, 75% correctly match the post quote. Thus, the quote rule is correct for about 95% of the total sample using pre and post quotes.⁹ The remaining 25% matched the wrong side of the post quote (a buyer initiated trade matched the post bid price); therefore, the quote rule is 80% correct with pre quotes and 95% correct with both pre and post quotes. We conservatively present our results with the 80% pre quote sample match because it is not clear why so many trades match the post quote – perhaps due to a systemic recording issue or 'slippage' of a few milliseconds in the processing time to record the trade.

Table 3 presents results for the 80% of trades (543,629) which correctly match the previous best quote's bid/ask according to the sell/buy initiator. Panel A details the tick rule correct buyer/seller classification percentage and the seven types of errors for the subsamples of trade direction, tick direction, and quote change. The first three columns are for trade direction – buyer, seller, and their average. The overall average correct allocation is 65.67% which is significantly less than previous research, specifically AF (74.4%) and Chakrabarty (75.6%). Of some concern is the asymmetry for trade direction between buyer (60.85%) and seller (70.34%) initiated trades. AF's results were more symmetrical - 75.1% (buyer) and 73.7% (seller) – with the seller being the poorer classification. Theissen was 72% (buyer) and 72.5% (seller), while Odders-White was 80% (buyer) and (77%) seller (calculated from her Table 2). The magnitude

⁹ The tick rule's average accuracy hardly changes (66.5%) for this subsample (trades matched to their post quote), compared to 65.7% for the main sample.

of the errors 1 and 2 were about 7% larger than AF – our error 1 was 8% (7.7%) for buyer (seller) while AF was 1.2% (1.7%) respectively and our error 2 was 7.49% (7.22%) for buyer (seller) while AF was 0.6% (0.7%) respectively. However AF type 3 error was 5% larger, type 5 was 7% larger, and type 6 was 2.5% larger. Our largest error was 16.34% for ‘other’ buyer initiated errors, while ‘other’ seller initiated errors were only 7.5% which explains the buyer/seller asymmetry in the correct classifications above. The ‘other’ errors here are multi zero tick situations (3 to 6 consecutive equal trade prices) which include all of the error types; however, it is not clear why buyer initiated trades are more adversely affected. Note that the reverse tick test did even worse with an average correct percentage of less than 50% with similar asymmetry between buyer and seller initiated trades. Therefore, it must be the case that quote revisions tend to follow the direction of order imbalance for price continuations (a current buy at the ask when the quote increases after a previous buy – see Lee and Ready (1991)). Consequently, we do not present the reverse tick test results because they are dominated by the tick rule.

[Insert Table 3 about here]

The middle three columns (Panel A of Table 3) provide the average correct percentage classification for the tick direction subsamples - down, zero, and up. The interesting result for tick direction is the expected zero tick lower accuracy of 61.19% (about 49.3% of the sample) compared to the uptick (69.74%) and downtick (70.31%) success rates. The major culprit appears to be the multi zero ticks termed as ‘other’ errors at 24.04% which is about 9% higher than the type 1 errors (type 3,4,5, and 6 errors total about the same as the type 2 errors). It is worth noting the symmetry for up and downticks for the correct classification (about 70%) and the type 1 and 2 errors (about 15% each). AF zero tick accuracy was 68.5% (about 75% of their sample) while the up (down) tick accuracy was 90.6% (92.8%) – about 21% higher than our up/downtick forex result. It is interesting that Chakrabarty reports 75% accuracy for zero ticks (72.4% of his sample) and 76% accuracy for up and down ticks – very little difference. Most previous research shows much lower accuracy for zero tick trades compared to up and downtick trades – the Appendix shows a range of 13% to 23% in four other papers.

The last four columns of Panel A in Table 3 report percentage accuracy and errors when the current mid quote has changed from the previous trade’s mid quote (about 60% of the sample). There is very little asymmetry between buyer/seller when the quote changes (62.18% versus 65.35%) but, remarkably, it increases substantially if there is no change in the mid quote – 58.81% (buyer) and 77.73% (seller), where ‘other’ errors is 31.3% (buyer) and 12.65% (seller). It would seem that multi zero ticks heavily affect buyer side accuracy when the mid quote does not change (about 40% of the sample). AF have conflicting results as their ‘quote change’ shows much greater asymmetry – 61.5% (buyer) and 53.9% (seller) – with the seller being less

accurate. It is difficult to explain such diametrically opposed results between the two papers, although we note that AF's 'quote change' subsample is only 7.2% of their trades compared to our 60%. Indeed, it is quite surprising that 92.8% of AF's 4,022,339 trades had no change in the mid quote. As mentioned at the end of the previous section, AF expected poorer tick classification when the mid quote changes because the tick rule biases classification in the direction of a trending market (a rising market is more likely to be labelled a buy). In our data the 'no quote change' average of buyer and seller is 68.27% compared to the 'quote change' average of 63.77%, so there is a slightly poorer average misclassification by 4.5% for the quote change forex subsample. So the direction, at least, is consistent with AF; however, their average difference is much larger - almost 18% (76% no quote change versus 58% quote change). We pursue this decrease of asymmetry for quote changes in Panel B.

To summarize Panel A of Table 3 with respect to trade direction, buyer initiated trades are more likely to be misclassified by 9.49% relative to seller initiated, due to multi zero tick errors. With respect to tick direction, zero tick trades are harder to classify correctly as they are about 8.84% on average less accurate than up or downticks. Lastly, multi zero tick errors increase the asymmetry for buyer misclassification relative to seller from a 9.49% difference to 18.92% when the mid quote does not change. In contrast, AF find greater overall tick test accuracy with little asymmetry for trade direction – the seller is 2% less accurate. They do note significant asymmetry when the mid quote changes from the previous trade's mid quote as the seller is 7.5% less accurate which they explain by the short sale constraint that prevents a short sale on down ticks in equity markets. However, in the forex market there is no such constraint; furthermore, we find the buyer side is much less accurate when there is no mid quote change. With respect to tick direction, correct zero tick rule classification is 61.2% (49.3% of our sample), compared to 68.5% in AF (75% of their sample) and 75.1% in Chakrabarty (72.4% of their sample). The difference for correct tick rule classification between zero ticks and up/down ticks is about 21% in AF compared to 9% in our Panel A (only 1% in Chakrabarty). Clearly, forex and Australian equity have completely different behavior notwithstanding their similarities – both are ECNs, automatic order match systems with all trades at quotes, full transparency, and no special privilege/advantage for intermediaries.

Since the trade direction asymmetry between buyer/seller decreased from roughly 10% for the main sample to 3% for the quote change columns (compared to 18% for no quote change) in Panel A of Table 3, we investigated further by splitting the quote change sample into upward and downward subsamples - the first 4 columns of Panel B. Note the up and down quote change columns have extreme results for correct classification – 95.82% (buyer, up) and 97.05% (seller, down). Recall that the quote change for these columns is the current trade's mid quote relative to the previous trade's mid quote, so the up (down) change has a higher (lower) mid

quote. Since the errors illustrated in Figure 1 are based on tick changes (not mid quote changes), we introduce Figure 2 to explain these extreme results.

Panel A of Figure 2 presents various scenarios for the current trade (t) in order to explain the 95.82% success rate for buyer initiated trades for the upward mid quote change. In all 7 scenarios the mid quote is higher than the previous mid quote and the current buy (at the bid) is an uptick for all except for scenarios 6 and 7 – both denoted sells by the tick rule. Therefore, the tick rule is incorrect for only scenarios 6 and 7, which are both very low probability scenarios as 6 is a zero spread (bid = ask) and 7 is a smaller spread than the previous (t-1) spread (since the average spread is about 1 bp, the current (t) spread would have to be half a bp with a higher bid and slightly lower ask). Note that the previous trade (t-1) is a buy in Panel A and that if it had been a sell at the bid, then the tick test would not make any errors as all scenarios would be upticks. Therefore, the 95.82% (correct buyer initiated) is so high because the tick rule is correct in all scenarios for a (t-1) sell and correct in scenarios 1, 2, 3, 4, and 5 for a (t-1) buy. The tick rule is only wrong for the two very low probability scenarios 6 and 7.

[Insert Figure 2 about here]

The seller initiated trades are only 28.77% correct when the mid quote increases. One wonders why it is not lower, say about 5%, in order to be symmetrical with the buyer column. To explain this, see Panel B of Figure 2, where the only change relative to Panel A is that the current trade (t) is seller initiated (it is at the bid for the 7 scenarios). When the previous trade (t-1) is a buy, scenarios 1, 2, 3, 6, and 7 are down ticks and are correctly classified as sells by the tick rule, but scenarios 4 and 5 are upticks and are incorrectly classified as buys. Now change the previous trade (t-1) to a sell at the bid, and only scenario 1 (a downtick) is correctly classified as a sell - all the other scenarios are up ticks and incorrectly classified as buys. As a result, there are more current trade scenarios on the sell side (Panel B) that are correctly classified than there are incorrectly classified for the buy side (Panel A), which explains why the 28.77% is not lower. A similar analysis for the downward changes explains the correct classifications for buyer (24.92%) and seller (97.05%).

Lastly, it is the volume weighted average of 95.82% (buyer, up) with 24.92% (buyer, down) in Panel B that yields the buyer quote change of 62.18% in Panel A. Apparently this average figure is deceptive as it is composed of two extremes – one overwhelmingly right and the other quite wrong (each extreme contains about 15% of our sample). This has the potential to severely affect related research sampling from the buyer side of down quote changes because signed trades in this category are less than 25% correct. Note the large error rates of about 27% (26%) for error 1 (2) respectively for buyer (down quote) which explains the majority of this misclassification (the rest are multi zero tick errors, and zero tick errors 3 and 4). This is

expected as falling mid quotes tend to decrease asks which increases the likelihood of a downtick, thereby causing the tick rule to misclassify the current buy. A similar situation exists for the seller quote change success rate of 65.35% which is the average of 28.77% (seller, up) and 97.05% (seller, down). Again, we see that the majority of the error in the seller side of up changes comes from error 1 (27%) and 2 (26%). This is expected as rising mid quotes tend to increase bids, which increases the likelihood of an uptick, thereby causing the tick rule to misclassify the current sell.

Panel B of Table 3 also presents subsamples for the daily movement of the mid quote – upward (downward) if the mid quote increased (decreased) by the end of the day compared to the previous day and zero otherwise. The asymmetry between buyer/seller first noticed in the trade direction of Table 3 persists for up, down, and zero movement days. Errors are remarkably consistent as well across all three scenarios with multi zero ticks being the largest category for buyer initiated trades.

We also examine the above tick test results over time by month in Figure 3 for the main sample that matches a trade to its previous quote. Panel A gives the monthly behavior over 2 years for correct buyer/seller classification, while the seven buyer and seller errors are in panels B and C respectively. Panel A shows fairly consistent buyer/seller asymmetry around 60/70% for the buyer/seller on the left axis with respect to the months in 2005/6. The difference, on the right axis, is within 4% of the average of 9.49% (70.34% (seller) minus 60.85% (buyer) from Table 3). The summer months have the lowest asymmetry (about 6% in August 2005 and June 2006) while the winter months have the highest (14% in December 2005 and 13.3% in February 2006). However, Panel B has considerable variability for multi zero tick errors for buyer initiated trades averaging about 16% with a range between 23% and 11% with a slight downward trend. Errors 1 and 2 are relatively stable within a range of 5% to 10% with a weak upward trend over the 2 years. The zero tick errors are low and stable over time. Panel D shows that monthly buyer trades are roughly equal to seller trades with a range of 7,000 to 17,000 trades/month with no discernible trend. The correct tick test average classification in Panel E is fairly stable until the last quarter of 2006 when the correct percentage on downticks declined substantially below the uptick line. This was offset by an increase in the zero tick correct percentage in late 2006 while the uptick result remained stable. Panel F compares the current trade's mid quote to the previous trade's mid quote for the correct tick test percentage over both years. The 'quote change' means the current and previous trades' mid quotes are not equal and we notice the correct seller percentage is always greater than the buyer by 2% to 6%. Recall from Table 3, Panel A that this average was 62.18% (buyer) and 65.35% (seller). However, when the current and previous mid quotes are equal (no quote change) the seller trades are more accurately classified than buyer by about 18%. Seller correct percentages are fairly stable around 74% to 82%, but buyer correct percentages are quite variable between 52% and 65%. To summarize,

the trends over time generally support the averages, but there is enough monthly variation to indicate that other factors are present. Nevertheless, the asymmetry between buyer and seller initiated trades is consistent across time, and seems to stem from multi zero tick errors that are most evident when there is no change between the current and previous trade's mid quote.

[Insert Figure 3 about here]

8. Conclusion

This paper is the first to examine the signing success of the tick rule in the forex market. The buyer/seller initiated forex trade data and their quotes for 2005/6 is a unique opportunity to definitively test the two major signing algorithms used in microstructure research - the tick test and the quote rule. Although previous research has used proprietary data to infer trade initiation, it has usually been with some error. The two main methods are the latest time stamp method of Odders-White (2000), while the other method requires counterparty identities with the assumption of passive intermediaries (as in Finucane (2000)). Both of these authors could not sign about a quarter of their data due to various issues such as missing time stamps or unassignable trades between intermediaries. In addition, both these methods must still match trades to quotes; this is complicated by various delays on different exchanges in recording trades with respect to quotes, which leads to potential errors. As a result, the availability of signed forex trades is a rare opportunity to reduce the complications and errors previously encountered.

Signed trades are important for microstructure research in such areas as PIN (probability of informed trading), effective spreads, price impact, market quality, short sales, and inventory theories. As noted by Theissen (2001), the accuracy of signing algorithms is critical. Boehmer et al (2007) and others note biases created by inaccurate signed trades, while Tangaard (2003) states that corrections or adjustments are not made even though most studies recognize the existence of signing errors. Perlin et al (2011) provides a stochastic model of the tick test that could be used to correct for errors, and we noted that the forex inventory effect of Lyons (1995) was not found by Bjonnes and Rime (2005) who used the Odders-White latest time stamp to sign forex trades instead of the tick test used by Lyons.

The Hotspot forex data on the USD/Euro comes from a fully computer automated order matching system in which all participants – including intermediaries – have no information advantage but have complete access to the order book. All trades occur on quotes, so we cannot directly examine LR and EMO and their variations, which must classify trades inside and outside the quotes. However, these methods do use a combination of the quote rule and the tick test, so our results have some bearing on their potential success in forex research. We

frequently compare our results to AF (Aitken and Frino (1996)) because their two years of data on Australian equity also comes from a similar 'trade at quote' automated transparent order book system.

In particular, we find the quote rule is 80% accurate in matching trades to the previous best quote while the tick test is only 66% correct on average. However this average disguises a 9.49% asymmetry between buyer (60.85%) and seller (70.34%) initiated trades which persists throughout our two year study period. This asymmetry is strongest for small spreads (1 bp or less which is 73.8% of the sample) and periods of fast trading (less than 3 seconds between trades which is 45.5% of the sample). Interestingly, this 9.49% average asymmetry increased to 18.92% for trades whose mid quote did not change from the previous trade's mid quote (about 40% of our sample). The asymmetry comes from multi zero tick errors which are much higher for buyer initiated trades. It is noteworthy that AF did not find asymmetry for their 'no quote change' subsample which was 92.8% of their total trades. When mid quotes differ from the previous trade's mid quote, the correct tick test percentage was 62.18% (buyer) and 65.35% (seller), while AF found 61.5% (buyer) and 53.9% (seller). In other words, AF found severe asymmetry for the mid quote change where we did not and their seller side was the poorer performer while ours was the buyer side. To investigate these contrary results we looked at the up and down change subsamples of the mid quote change scenario. We find the up change correct tick test percentages are 95.82% (buyer) and 28.77% (seller), while down changes are 24.92% (buyer) and 97.05% (seller). We explain these large asymmetries through an examination of the various tick scenarios for each subsample. We also note that microstructure research which primarily samples from buyer (seller) side down (up) quote changes has a terrible error rate of roughly 75% (71%) respectively. However, it is not clear why our overall sample has significantly more buy side signing errors even though we identify specific subsamples that have larger asymmetry.

In conclusion, we recommend the quote rule for signing forex trades with matching to the previous best quote. The tick test has significant errors and the reverse tick test is even worse. Clearly, some adjustment or corrections should be made to tick rule signed forex trades.

References

- Aitken, M., Frino, A., 1996. The accuracy of the Tick Test : Evidence From the Australian Stock Exchange. *Journal of Banking and Finance* 20, 1715-1729.
- Anand, A., Subrahmanyam, A., 2008. Information and the Intermediary: Are Market Intermediaries Informed Traders in Electronic Markets?. *Journal of Financial and Quantitative Analysis* 43, 1-28.
- Anand, A., Tanggaard, C., Weaver, D.G., 2009. Paying for Market Quality. *Journal of Financial and Quantitative Analysis* 44, 1427-1457.
- Angel, J., Harris, L., Spatt, C.S., 2010. Equity Trading in the 21st Century. Working Paper FBE 09-10.
- Asquith, P., Oman, R., Safaya, C., 2008. Short Sales and Trade Classification Algorithms. NBER Working Paper 14158.
- Berger, D.W., Chaboud, A.P., Chernenko, S.V., Howorka, E., Wright, J.H., 2008. Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data, *Journal of International Economics* 75, 93-109.
- Bessembinder, H., 2003. Issues in Assessing Trade Execution Costs. *Journal of Financial Markets* 6, 233-257.
- Blume, M., Mackinlay, A., Terker, B., 1987. Order imbalances and stock price movements on October 19 and 20, 1987. *Journal of Finance*, 44, 827-848.
- Bjønnes, G.H., Rime, D., 2005. Dealer Behavior and Trading Systems in Foreign Exchange Markets. *Journal of Financial Economics* 75, 571-605.
- Boehmer, E., 2005. Dimensions of execution quality: Recent evidence for US equity markets. *Journal of Financial Economics* 78, 553-582.
- Boehmer, E., Grammig, J., Theissen, E., 2007. Estimating the probability of informed trading—does trade misclassification matter?. *Journal of Financial Markets* 10, 26-47.
- Chaboud, A., Chiquoine, B., Hjalmarsson, E., Vega, C., 2012. Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. FRB International Finance Discussion Paper 980.
- Chakrabarty, B., Li, B., Nguyen, V., Van Ness, R.A., 2007. Trade Classification Algorithms for Electronic Communications Network Trades. *Journal of Banking and Finance* 31, 3806-3821.
- Chakrabarty, B., Moulton, P.C., Shkilko, A., 2012. Short Sales, Long Sales and Lee-Ready Trade Classification Algorithm Revisited. *Journal of Financial Markets* 15, 467-491.
- Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order Imbalance, Liquidity, and Market returns. *Journal of Financial Economics* 65, 111-130.

- Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and Market Efficiency. *Journal of Financial Economics* 87, 249-268.
- Chung, D.Y., Hrazdil, K., 2010. Liquidity and market efficiency: A large sample study. *Journal of Banking and Finance*, 34, 2346-2357.
- Chung, D.Y., Hrazdil, K., 2012. Speed of Convergence to Market Efficiency: The Role of ECNs. *Journal of Empirical Finance* 19, 702-720.
- Ellis, K., Michaely, R., O'Hara, M., 2000. The Accuracy of Trade Classification Rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis* 35, 529-551.
- Erenburg, G., Lasser, D., 2009. Electronic order book and order submission choice around macroeconomic news. *Review of Financial Economics* 18, 172-182.
- Evans, M.D.D., Lyons, R.K., 2002. Order Flow and Exchange Rate Dynamics. *Journal of Political Economy* 110, 170-180.
- Evans, M.D.D., Lyons, R.K., 2008. How is Macro News Transmitted to Exchange Rates?. *Journal of Financial Economics* 88, 26-50.
- Finucane, T.J., 2000. A Direct Test of Methods for Inferring Trade Direction from Intra-Day Data. *Journal of Financial and Quantitative Analysis* 35, 553-576.
- Garvey, R., Wu, F., 2010. Speed ,distance, and electronic trading: New evidence on why location matters. *Journal of Financial Markets* 13, 367-396.
- Glosten, L.R., 1987. Components of the bid-ask spread and the statistical properties of transaction prices. *Journal of Finance*, 42, 1293-1307.
- Glosten, L.R., 1994. Is the Electronic Open Limit Order Book Inevitable? *Journal of Finance*, 49, 1127-1161.
- Hasbrouck, J., 1988. Trades, quotes, inventories, and information. *Journal of Financial Economics*, 22, 229-252.
- Hasbrouck, J., Saar, G., 2009. Technology and Liquidity Provision: The Blurring of Traditional Definitions. *Journal of Financial Markets* 12, 143-172.
- Hendershott, T., Moulton, P.C., 2011. Automation, Speed, and Stock Market Quality: The NYSE's Hybrid. *Journal of Financial Markets* 14, 568-604.
- Hendershott, T., Riordan, R., 2009. Algorithmic Trading and Information. Working Paper 09-08, NET Institute.
- Holden, C.W., Jacobsen, S., 2011. The Breakdown of Standard Microstructure Techniques: And What to Do About It. Working Paper, IU Kelley School of Business.

- Huang, R.D., 2002. The Quality of ECN and Nasdaq Market Maker Quotes. *The Journal of Finance* 53, 1285-1319.
- Huh, S.W., 2013, Price Impact and Asset Pricing. *Journal of Financial Markets*, In Press.
- Jain, P.K., 2005. Financial Market Design and the Equity Floor Premium: Electronic versus Floor Trading. *Journal of Finance*, 60, 2955-2985.
- Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-1356.
- Lee, C.M.C., Radhakrishna, B., 2000. Inferring Investor Behavior: Evidence from TORQ Data. *Journal of Financial Markets* 3, 83-111.
- Lee, C.M.C., Ready, M.J., 1991. Inferring Trade Direction from Intraday Data, *The Journal of Finance* 46, 733-746.
- Lu, Y., Wei, Y., 2009. Classification of Trade Direction for An Equity Market with Price Limit and Order Match: Evidence From the Taiwan Stock Market. *Investment Management and Financial Innovation* 6, 135-147.
- Lyons, R.K., 1995, Tests of Microstructural Hypothesis in the Foreign Exchange Market. *Journal of Financial Economics* 39, 321-351.
- Odders-White, E.R., 2000. On the Occurrence and Consequences of Inaccurate Trade Classification. *Journal of Financial Markets* 3, 259-286.
- Perlin, M., Brooks, C., Dufour, A., 2011. On the Performance of the Tick Test. Working paper, UFRGS.
- Petersen, M., and Fialkowski, D., 1994. Posted versus effective spreads: Good prices or bad quotes?. *Journal of Financial Economics*, 35, 269-262.
- Piowar, M., Wei, Li., 2006. The Sensitivity of Effective Spread Estimates to Trade-Quote Matching Algorithms. *Electronic Markets*, 16, 112 -129.
- Riordan, R., Storkenmaier, A., 2012. Latency, Liquidity and Price Discovery. *Journal of Financial Markets*, 15, 416-437
- Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39, 1127-1139.
- Stoll, H., 2006. Electronic trading in stock markets. *Journal of Economic Perspectives*, 20, 153-174.
- Stoll, H., Schenzler, C., 2006. Trades outside the quotes: Reporting delay, trading option, or trade size? *Journal of Financial Economics* 79, 615-653.
- Tinggaard, C., 2003. Errors in Trade Classification: Consequences and Remedies. Working Papers 03-6, AU, Aarhus School of Business.

Tay, A., Ting, C., Tse, Y.K., Warachka, M., 2009. Using High-Frequency Transaction Data to Estimate the Probability of Informed Trading. *Journal of Financial Econometrics* 7, 288-311.

Theissen, E., 2001. A Test of the Accuracy of the Lee-Ready Trade Classification Algorithm. *Journal of International Financial Markets, Institutions and Money*, 11, 147-165.

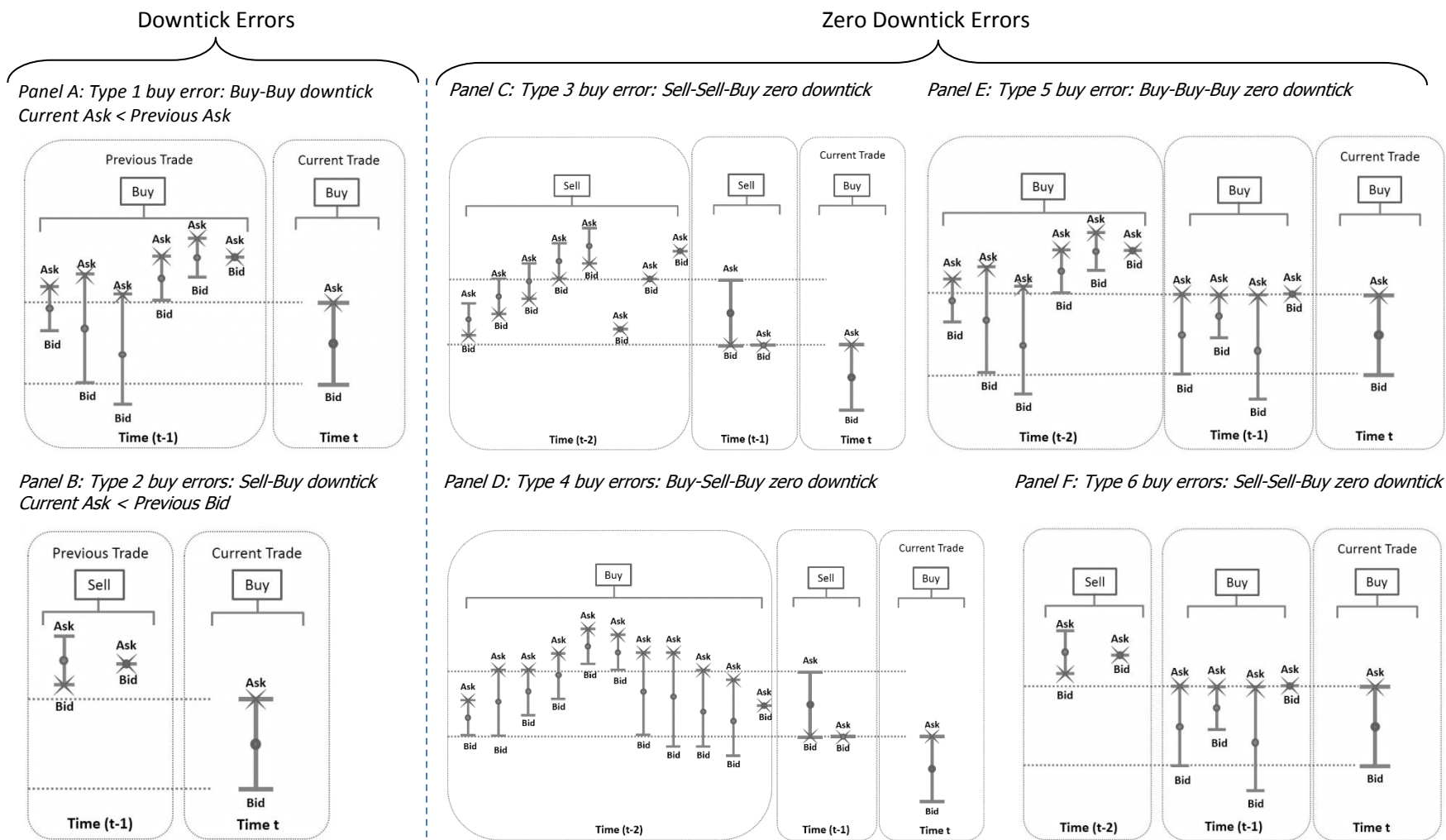
Vergote, O., 2005. How to Match Trades and Quotes for NYSE Stocks, Working Paper.

Appendix: Literature Summary

Authors	Data			Matching Procedure	Algorithm Average Success Rate (%)				Special Notes (% accuracy of specific subsamples)
	Type/Period	Size	Composition		Tick Test	Quote Rule no midquote	LR	EMO	
Lee & Radhakrishna (2000)	Equity (Specialist) NYSE/TORQ 3 months (1990/1)	144 firms 129,700 trades (29% of original sample)	77% at quote 19% midquote 5% inside quote	Order/ trader type	-	96 [84]*	93	-	Midquote 76 (zero tick 61)
Odders-White (2000)	Equity (Specialist) NYSE/TORQ 3 months (1990/1)	144 firms 318,364 trades (75% of original sample)	81% at quote 16% midquote 3% inside quote	Latest time stamp	79	89[75]*	85	-	Midquote 63 (zero tick 60, nonzero tick 77)
Finucane (2000)	Equity (Specialist) NYSE/TORQ 3 months (1990/1)	144 firms 285,983 trades (75% of original sample)	76% at quote 20% midquote 3.5% inside quote	Order/ trader type	83	89 [71]* (inferred from table 2)	84	-	TT/QR zero tick 77/79, nonzero tick 96/96; at quotes 88/89; Midquote 66
Ellis et al. (2000)	Equity (Dealer) NASDAQ/TAQ 3-12 months (1996/7)	313 firms 1.8 million trades (75% of original sample)	70% at quote 10% midquote 15% inside quote	Order/ trader type	78	83 [76]*	81	82	TT/QR at quotes 83/89 Midquote 61
Theissen (2001)	Equity (Specialist) DAX(Germany) 21 days (1996)	15 stocks, 9,449 trades (55% of trade volume with Makler)	-	Order/ trader type + buy/sell by Makler	72	75	73	-	Midquote zero tick 53, nonzero tick 73; TT(full sample) zero tick 60, nonzero tick 73
Aitken & Frino (1996)	Equity (ECN) ASX(Australia) 2 years (1992/4)	4 million trades (most of original sample)	100% at quote (exclude crosses & off market trades)	Latest time stamp	74	-	-	-	TT zero tick 69, nonzero tick 92
Chakrabarty et al. (2007)	Equity (ECN) NASDAQ/ITCH/TAQ 3 months,2005	750 firms 17.4 million trades (price/sh ≥5)	58% at quote 19% midquote 14% inside quote	Latest time stamp plus buy/sell from ITCH	75	76 [62]* (inferred from table 1)	74	76	TT/QR at quotes 79/80 Midquote 69 TT/QR zero tick 75/73, nonzero tick 76/77
Savickas & Wilson (2003)	Options (Market Maker, Specialist) CBOE/OPRA 1 year (1995)	826 stocks, 90 million options, 1.4 million trades(73% of original sample)	67% at quote 11% midquote 15% inside quote	Order/trader type +buy/sell from ITCH	59	83 [79]*	80	77	TT/QR at quote 63/84 Midquote 48**
Perlin et al. (2011)	Equity (ECN) BOVESPA(Brazil) 1 year (2009)	15 stocks 13 million trades (100% of original sample)	100% at quote	Latest time stamp	72	-	-	-	-
Lu & Wei (2009)	Equity (ECN) TWSE(Taiwan) 6 months (2006)	684 stocks 17.2 million trades	94% at quote 3% midquote	Modified Quote Rule (RLR) for one sided trades	74	93	97	95	Order matching system: RLR can classify all trades and is assumed to be correct

Notes: * counts midquote trades as errors. ** Index options, complex and reverse quote trades poorly classified by all algorithms and TT has higher proportion of these trades.

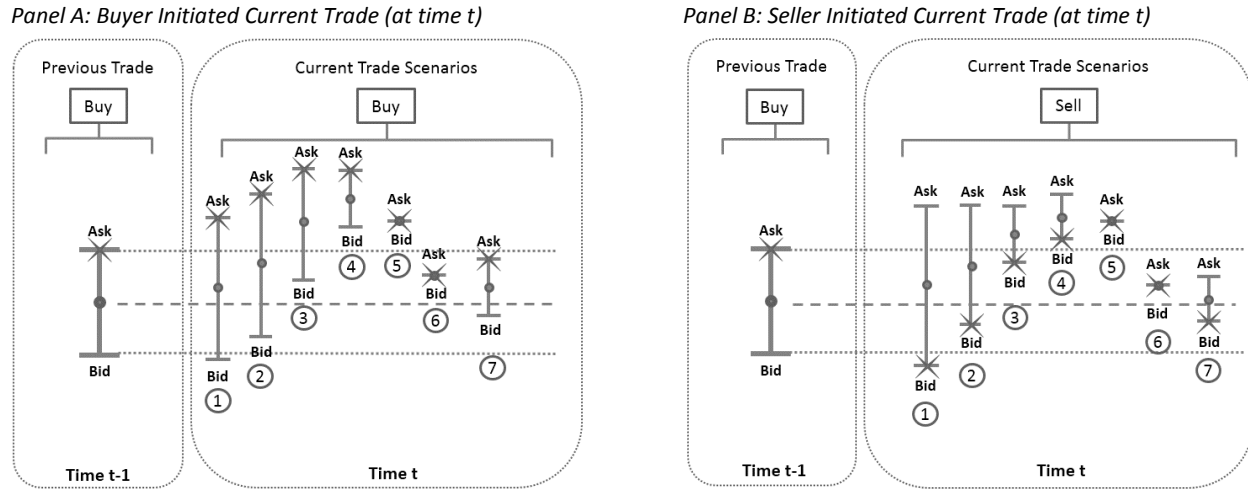
Figure 1: BUYER INITIATED TICK RULE ERRORS



✕ Stands for transaction price and ● the mid quote price. The above six errors coincide with the Aitken and Frino (1994) classifications. They illustrate tick test error when assigning a buy to the current trade on a downtick or zero downtick when the current quote (panels A and B) falls relative to the previous trade's quote, or for zero ticks, when the previous trade's quote falls relative to its prior trade's quote (panels C, D, E, and F). In all cases the tick rule assigns a sell because of the down tick (or zero downtick), yet the trade is a buy at the ask.

Figure 2: Upward Mid Quote Change

The current trade's mid quotes exceeds the previous trades' mid quote:

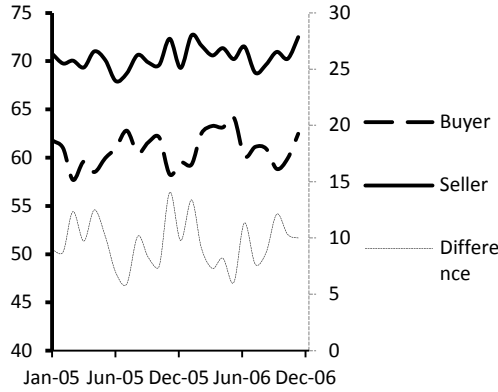


✕ Stands for transaction price and ● the mid quote price. The tick rule incorrectly assigns a 'seller' initiated trade at the current time (t) in scenarios ⑥ and ⑦ of panel A because they are downticks. In panel B the tick rule incorrectly assigns a 'buyer' initiated trade at the current time (t) in scenarios ④ and ⑤ because they are upticks.

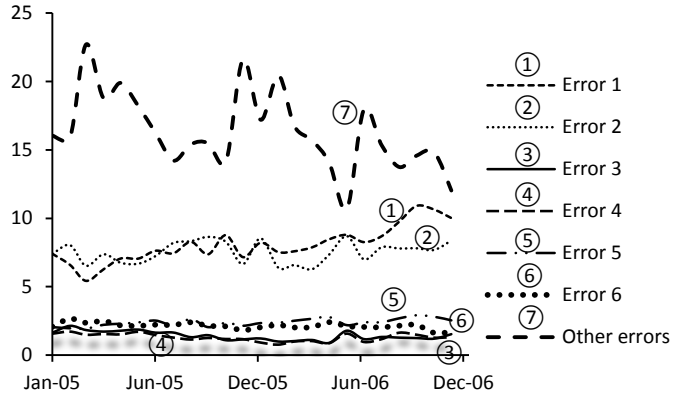
If the previous trade (at (t-1)) in panel A is a sell at the bid then then the tick test makes no errors (all seven scenarios at t are upticks which are buys according to the tick rule). For panel B, if the previous trade (at (t-1)) is a sell at the bid, then the tick rule incorrectly assigns buy in scenarios ②, ③, ④, ⑤, ⑥, ⑦ (all these scenarios are upticks).

Figure 3: Monthly Behaviors of Selected Statistics

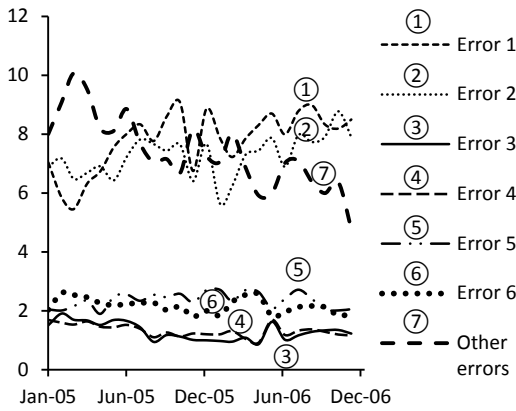
Panel A: Monthly Correct Tick Test Classification (%)



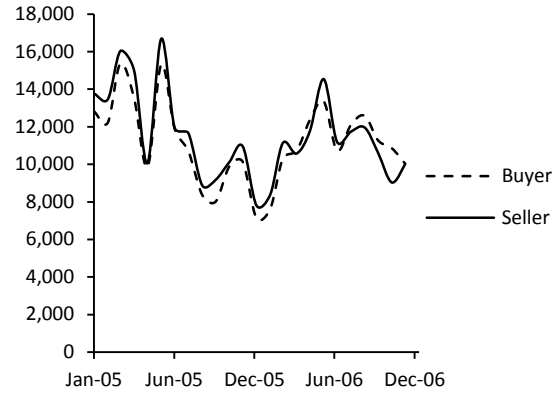
Panel B: Monthly Buyer Errors (%)



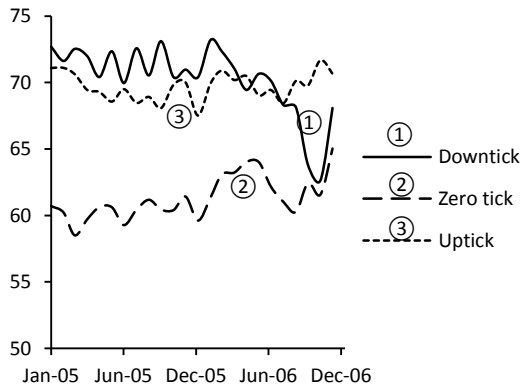
Panel C: Monthly Seller Errors (%)



Panel D: Monthly Trades Initiated by Buyer/Seller



Panel E: Tick Direction (%)



Panel F: Quote Change/No Quote Change (%)

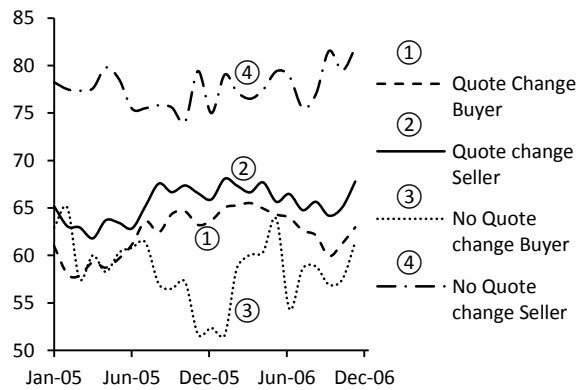


TABLE 1

TABLE 1: Tick Test Accuracy by Spread Size (Ask-Bid)

Quote spread (Ask-Bid) is categorized by size in basis points (1 bp = 0.01%) for buyer and seller initiated trades in 2005/6 for the sample matched to the previous best quotes (543,629 trades)

Spread Size (bp)	Total Sample	Correct classification %	
	%	Buyer	Seller
0	3.70	43.14	52.16
(0 , 0.5)	23.30	52.30	66.04
[0.5 , 1)	46.80	61.19	67.34
[1 , 1.5)	12.11	70.48	70.51
[1.5 , 2)	11.53	72.61	72.72
[2 , 2.5)	2.55	77.36	75.67

Mean spread is 1.09 with a standard deviation of 0.9.

TABLE 2: Time between Trades

The time between consecutive trades in seconds is categorized for buyer and seller initiated trades in 2005/6 for the sample matched to the previous best quote (543,629 trades).

Time between trades: (seconds)	Total Sample	Correct classification %	
	%	Buyer	Seller
0.01<d<=0.05	0.16	2.15	2.99
0.05<d<=0.10	8.99	49.77	75.95
0.10<d<=0.15	9.40	58.91	76.68
0.15<d<=0.20	4.16	61.03	75.13
0.20<d<=1	12.97	58.42	73.02
1<d<=2	5.98	60.25	71.88
2<d<=3	3.84	63.20	72.43
3<d<=4	2.90	63.80	70.34
4<d<=10	9.94	63.52	70.28
10<d<=60	21.82	63.97	67.13
60<d<=600	16.79	62.13	63.49
600<d	2.95	60.51	59.28

Mean time between trades is 18.63 seconds with a standard deviation of 40.67 seconds. Mean time between a trade and its previous quote is 1.13 seconds with a standard deviation of 46.45 seconds.

TABLE 3: Tick Test Accuracy and Errors for 2005/6

Panel A: The sample includes all trades that correctly matched the previous best quoted (almost 80%). Trade direction refers to buyer or seller initiated trades. Tick direction refers to the location of the current trade to the previous trade. Quote change means the current trade's mid quote differs from the previous trade's mid quote. Errors 1 and 2 are up/down tick errors, 3 to 6 are zero up/down tick errors and other errors are multi-zero tick errors.

%	Trade direction			Tick direction			Quote change		No Quote change	
	Buyer	Seller	Average	Downtick	Zero tick	Uptick	Buyer	Seller	Buyer	Seller
Correct	60.85	70.34	65.67	70.31	61.19	69.74	62.18	65.35	58.81	77.73
Error 1	8.00	7.70	7.85	15.33	0.00	15.62	13.00	12.69	0.35	0.31
Error 2	7.49	7.22	7.35	14.36	0.00	14.64	12.39	12.09	0.00	0.00
Error 3	1.45	1.34	1.40	0.00	2.84	0.00	2.37	2.23	0.05	0.03
Error 4	1.33	1.36	1.35	0.00	2.73	0.00	2.18	2.25	0.04	0.03
Error 5	2.39	2.34	2.36	0.00	4.80	0.00	0.71	0.69	4.94	4.79
Error 6	2.16	2.19	2.17	0.00	4.41	0.00	0.61	0.65	4.51	4.47
Other errors	16.34	7.51	11.85	0.00	24.04	0.00	6.55	4.05	31.30	12.65
Tot trades	267,122	276,507	543,629	139,352	267,944	136,333	161,432	165,046	105,690	111,461

Panel B: The quote change columns in panel A are split into upward (downward) change subsamples where the current trade's mid quote is greater (smaller) than the previous trade's mid quote respectively. The remaining six columns refer to the daily movement of the mid quote between the start and end of each day.

%	Quote change				Daily movement					
	Upward change		Downward change		Upward movement		Downward movement		Zero movement	
	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller	Buyer	Seller
Correct	95.82	28.77	24.92	97.05	60.09	70.34	61.51	70.34	60.72	70.32
Error 1	0.23	27.10	27.15	0.18	8.04	7.83	7.95	7.60	8.13	7.48
Error 2	0.00	26.04	26.12	0.00	7.45	7.22	7.48	7.25	8.06	6.85
Error 3	0.00	4.80	5.00	0.00	1.44	1.32	1.45	1.35	1.70	1.60
Error 4	0.00	4.84	4.60	0.00	1.29	1.34	1.35	1.36	1.53	1.56
Error 5	0.75	0.65	0.68	0.73	2.42	2.38	2.34	2.31	2.60	2.38
Error 6	0.65	0.60	0.57	0.69	2.16	2.15	2.15	2.23	2.26	2.05
Other errors	2.55	7.20	10.98	1.34	17.13	7.42	15.75	7.57	15.00	7.75
Tot trades	84,840	76,637	76,592	88,409	119,783	123,239	137,572	142,660	9,767	10,608