On the Performance of the Tick Test

Marcelo Perlin (msperlin@ea.ufrgs.br)*
Chris Brooks (c.brooks@icmacentre.ac.uk) †
Alfonso Dufour (a.dufour@icmacentre.ac.uk)‡

First Version: July 19, 2011
This version: November 23, 2011

EFM Classification Code: 330, 360

Working Paper

Abstract

In the present paper, we investigate the accuracy of the tick test from an analytical perspective by providing a closed formula for the performance of the prediction algorithm. This formula takes as inputs the spread of the traded asset, the volatility of the innovations, and the probability of news, among other parameters, and it outputs the percentage of times that the tick test will make correct predictions regarding the sign of a trade. Further analysis shows that by imposing restrictions on the underlying microstructure model, the formula for the tick test performance is related to simple statistics from a vector of trade price differences. This means that, without the need for quote data (or the real sign of the trades), the formula can assess the percentage of cases for which the tick test will make correct predictions. Using tick data for fifteen heavily traded stocks in the Brazilian equity market, we are able to compare the values from the analytical formula against the empirical performance of the tick test, showing that the formula is quite realistic in assessing the accuracy of the prediction algorithm.

*UFRGS - Federal University of Rio Grande do Sul (Porto Alegre, Brazil)
†ICMA Centre - University of Reading (Whiteknights Reading, RG6 6BA, United Kingdom)
‡ICMA Centre - University of Reading (Whiteknights Reading, RG6 6BA, United Kingdom)

1 Introduction

In the financial literature, the sign of a trade is important since it provides information on who in the trading process is the aggressor – that is, which side initiated the trade, the buyer or the seller. In other words, it gives information on who is the party demanding liquidity and who is the party offering it. Since liquidity is not free, the aggressor of a trade will pay for it in the form of a portion of the spread. The instant demand for liquidity has implications for microstructure theories. For instance, the proportion of buy and sells has an impact on the construction of a measure of the probability of informed trading (PIN)[1] and signed volume, which is typically used in the estimation of price impact models.[2]

The tick test is a simple algorithm which can be used to infer the sign of a trade when no quote data are available. While the origins of this method are still not clear, it was made popular in the work of Lee and Ready [1991]. According to the authors, its simple implementation and the fact that quote data are sometimes hard to obtain attracted the academic community, market regulators and traders to use it for inferring the trade direction from a sample of traded prices.

The research in the present paper is directly related to the predictive power of the tick test. Many previous studies have investigated the empirical performance of the algorithm, but none have understood it as a simple stochastic problem. In this paper, we take an analytical approach in this matter. Based on a microstructure model, a closed form solution for the performance of the algorithm is derived.

The paper makes several contributions to the literature. First, it is the only study to analyse the tick test as a stochastic problem and to show that its performance (number of correct predictions) can be represented as a simple mathematical formula. As far as we are aware, we are also the first to provide a formal argument on the use of the tick test by showing that it is bound to perform better than chance (50% correct predictions). Second, we show that, in the absence of quote data, the formula derived can still be used by assuming a particular reduction on the underlying microstructure model. This means that based only on data for trades, the researcher can still access the implied bias in the use of the tick test. This contributes to the literature (see Boehmer et al. [2006] and Tanggaard [2003]) which poins out the consequences of the missclassification of trade signs with respect to the construction of variables commonly used in microstructure research. The formula derived in the paper can be directly applied for the analysis of the resulting bias in the case of the tick test.

The remainder of the paper is organised as follows. First, we present a brief literature review on the subject of the tick test. In the second part of the paper, we develop the theory and show the formula derived for the performance of the algorithm. This is followed by an empirical examination of the accuracy of the analytical formula for trading data in the Brazilian equity market. The paper finishes with the usual concluding remarks.

[1]See Frank and Rindi [2009] for details.
[2]See Hasbrouck [1991] and Hasbrouck [2007].

2 Related Literature

The use of the tick test and other types of trade inference algorithms is popular in the academic literature. Just to cite a few, Lee [1993], Chordia et al. [2008], Chordia et al. [2005], and Chakravarty [2001], among many others, have used these methods to assess the sign of trades in their research.

The performance of the tick test algorithm (hereafter referred to as TT) has been extensively tested using empirical data, generally with good results. Lee and Ready [1991] conclude that the price-based algorithm presented "remarkably accurate" performance when classifying a sample of trades from the NYSE market. They also highlight an issue with the time stamps from the quotes and introduce an alternative methodology for predicting the sign of a trade, which is commonly known as the Lee and Ready algorithm (hereafter referenced as LR). This paper marked an important point in the literature as it was the first to formally study the inference of trade direction based on incomplete data.

A subsequent paper on the topic is by Aitken and Frino [1996]. Using data for the Australian stock Market, the authors find that the tick test produces approximately 74% correct predictions for the sign of a trade. The study also points out that periods with high volatility have a tendency to reduce the performance of the prediction algorithm. Following this study, we have the work of Theissen [2000]. Using data on the Frankfurt Stock Exchange, he finds that the accuracy of both algorithms (LR and TT) is comparable. The LR method produces 72.8% correct classifications while the tick test achieves 72.2% correct. He also shows the impact of the mis-classification of trades with respect to the estimation of the effective spread and and the PIN.[3]

Another empirical study on the tick test is due to Ellis et al. [2000]. The authors study the performance of distinct trade classification algorithms, including the quote rule, the LR rules and the tick test, for the NASDAQ market. This study also reports a positive performance of the trade inference algorithms for the sample data, which corroborates the results of Theissen [2000]. Further analysis undertaken in this paper also suggests that the performance of the classification rules decreases for large trades and also for periods of rapid trading (a small interval between previous trades and also for previous quote changes).

A more detailed analysis on the performance of the different trade sign forecasting methods is provided in Finucane [2000]. This study shows that trade size, spread and frequency of trades and quotes all affect the accuracy of prediction algorithms such as the LR and the tick test. The authors also report that, on average, the tick rule and the LR approach give very similar performance. Note that the work of Finucane [2000] already shows some of the findings presented in this paper, specifically that the spread will contribute in a positive way to the performance of the tick test (i.e., the higher the spread, the higher the accuracy of this particular method of prediction of the sign of a trade). However, we take the analysis a stage further and consider a wider range of variables.

In the same vein of the empirical analysis of trade inference algorithms, we have the work of Asquith et al. [2008]. This paper argues that the Lee and Ready

[3] The calculation of the PIN measure was introduced by Easley et al. [1996].

algorithm is most likely to fail when short sales are classified. The authors point out that short sales constitute a significant portion of the trading data for the US market, meaning that the estimates calculated based on the LR algorithm are very likely to have a significant bias.[4]

While many studies have investigated the empirical performance of the different trade inference algorithms, another interesting focus of research in this particular area of market microstruture theory is the bias resulting from the use of forecasted trade signs for the construction of different variables. This bias is a natural consequence of the fact that the forecasted trade signs have a degree of uncertainty, meaning that the predicted signs of trades are subject to errors (e.g., a buy misclassified as a sell). One of the studies on this topic is Boehmer et al. [2006], which shows that estimates of the PIN variable are downward biased in the presence of inaccurate classifications of trade signals. We also have the study of Tanggaard [2003]. In this paper, the author investigates the bias that trade classification errors can spur in the regression-type models typically used in market microstructure research. The author argues that this bias is probably even worse than the literature suggests, posing the question as to whether previous empirical results based on trade sign inference should be taken seriously.

The present paper extends some of this previous research by providing a simple way to assess the performance of the tick test. For example, Boehmer et al. [2006] and Tanggaard [2003] present a formal analysis on the bias from using trade inference methods in the construction of variables by using a measure for the probability of an incorrect inference for a trading sign. While the authors assume in their analysis that such a probability is given, the present paper can provide an easy way to estimate it from empirical data, at least for the case of the tick test.

Some of the empirical results already in the literature corroborate the findings in this study. As we will show later, two parameters that define the performance of the tick test are the spread and the volatility of price innovations. In the work of Aitken and Frino [1996], it is found that an increase in market volatility will reduce the performance of the trade inference algorithms. This is similar to the findings of Ellis et al. [2000], but the authors explicitly test for a relationship between trading frequency (which is related to incoming of news and volatility) and the performance of the tick test.

Our contribution to the literature lies in looking into the tick test algorithm from an analytical point of view. As we show, its performance has a simple closed formula, which, under certain restrictions, can be used even in the absence of quote data. With the use of the formula we derive, a researcher will be able to assess the implied performance of the tick test based only on the variance and autocovariance of trade price differences. In the next section, we present formal derivations of the results presented in this research.

[4]See also Odders-White [2000] and Peterson and Sirri [2003].

3 The Performance of the Tick Test

Consider a microstructure model comprising of the following set of equations:

$$\begin{aligned} m_t &= m_{t-1} + \Gamma_t \epsilon_t && (1) \\ P_t &= m_t + b_t \frac{S}{2} && (2) \\ \epsilon_t &\sim N(0, \sigma^2) && (3) \end{aligned}$$

$$b_t = \begin{cases} 1 & \text{if } ST_t = 1 \\ -1 & \text{if } ST_t = 2 \end{cases} \quad (4)$$

$$\Gamma_t = \begin{cases} 1 & \text{with probability } pNews \\ 0 & \text{with probability } (1 - pNews) \end{cases} \quad (5)$$

ST_t will follow a Markov chain, with transition probabilities given by the following matrix:

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \quad (6)$$

The specification in equations (1) to (4) represents a detailed microstructure model for the evolution of traded prices. The main innovations are due to the addition of uncertainty regarding the role of news and also to the representation of trades as a Markov chain process. Equation (1) defines m_t, which is the true (efficient) price of the asset and follows a random walk. It is straightforward to show that changes in m_t are not predictable by realising that $m_t - m_{t-1} = \Gamma \epsilon_t$, which is a censored Gaussian noise process with zero expectation. The financial intuition behind equation (1) is that the true price of the tradable asset is driven by the incoming of news with respect to the future cash flows of the financial instrument (stocks, bonds, etc). The shock of news on m_t is directly represented by ϵ_t. Given the symmetry of the Gaussian distribution, the probability of good news ($\epsilon_t > 0$) is equal to the probability of bad news ($\epsilon_t < 0$).

In our particular microstructure model, we can also see that the incoming of news is not guaranteed. There are moments in which there is news regarding the asset ($\Gamma_t = 1$), while there are other moments in which no news is available to the market ($\Gamma_t = 0$), implying that in these cases there will be a continuation in the efficient price, that is $m_t = m_{t-1}$. The incoming (or not) of news is defined by a probability $pNews$. As we will explain later in the paper, parameter $pNews$ plays a significant role in the performance of the tick test as it induces trade price changes equal to zero ($\Delta P_t = 0$), one of the triggers of a particular part of the tick test algorithm.

Equation (2) represents the traded price (P_t), which is a function of the efficient price plus a term for half the spread ($S/2$). The parameter b_t is the sign of the trade, taking the value 1 for a buy order and -1 for a sell. This parameter also has random behaviour where the occurrence of a buyer (seller) initiated order relates to a Markov chain, with transition probabilities given by (6). Representing trades as a Markov chain adds a time dependency to the process, meaning that the chance that a buy (sell) order occurs at any point in time is related to the sign of the previous trade. This autoregressive property is standard in empirical trading data, where the sign of the trades presents a positive and significant first order autocorrelation. The usual explanation is that traders split a big trading order in smaller ones, therefore causing a positive time dependency of the trade signs. So, the justification for using a Markov chain as the process for the trades is that it can mimic the time dependency usually found in empirical data.

Notice that the transition probabilities of state 1 ($ST_t = 1$,$b_t = 1$) and state 2 ($ST_t = 2$, $b_t = -1$) are equal (see (6)). This arises as a result of a simple assumption that the unconditional probability of a buy order is equal to the unconditional probability of a sell order.[5] The equality of the transition probabilities in (6) is a direct consequence of this assumption, which is strongly corroborated by the empirical data. The use of this simplification of the trade process reduces the number of parameters in the model without loss of information, therefore facilitating the derivations.

Examining the trade equation, we can see that when a buy order arrives at the market, the traded price will be given by $P_t = m_t + \frac{S}{2}$ and when a sell order arrives at the market, the price of the trade will be $P_t = m_t - \frac{S}{2}$. Notice that the aggressor side of the trade will always pay half the spread. This is the price for demanding liquidity. The system of equations given above represent a market with a constant spread S and no market maker (or a market maker with no sensitivity to inventory or asymmetric information).

Using the same notation as above, we now formalise the tick test in a simple set of rules. Consider P_t a trade price, ΔP_t its first difference, and $\hat{b_t}$ the predicted trade direction for time t. The rules from the tick test say:

- if $\Delta P_t > 0$ set $\hat{b_t} = 1$
- if $\Delta P_t < 0$ set $\hat{b_t} = -1$
- if $\Delta P_t = 0$ set $\hat{b_t} = \hat{b}_{t-1}$

As one can see, the intuition behind the first two rules of the algorithm is that a buy at time t is likely to increase the traded price from the previous trade. So, by inverting the logic, one can say that a buy (sell) order is most likely when the current trade price is higher (lower) than the previous one. For the cases where the change in the trade price is equal to zero, the intuition is to use a price continuation, meaning that the current forecast ($\hat{b_t}$) is set equal to the previous forecast ($\hat{b}_{t-1}$). Notice that this creates a recursive dynamic and, as long as the first element of the price difference vector is non-zero, the algorithm will always be able to build trade sign forecasts based on the historical data. For the cases of consecutive zero price movements ($\Delta P_t = 0$, $\Delta P_{t-1} = 0$, ... , $\Delta P_{t-k} = 0$), the

[5]By following a generic transition matrix, we can show that the unconditional probability of state 1 is given by $Pr(ST_t = 1) = \frac{1}{1-\frac{1-p_1}{p_2-1}}$. Therefore, by equating $Pr(ST_t = 1) = Pr(ST_t = 2) = 0.5$, we get the result that $p_1 = p_2$.

forecast of the trade sign will be defined using the forecast from a distant time period.

Now, with the help of a simple example, we will show the intuition which drives the analytical examination of the tick test. Consider a scenario where the interest is in forecasting, within a particular time period, the days with rain. Variable y_t will take the value 1 if there is rain on day t and zero otherwise. The researcher is testing the hypothesis that y_t is a function of a second variable, humidity (x_t). For some reason, he/she believes that a certain level of humidity on the previous day has forecasting power for the likelihood of rain the next day. Assume that both variables are easily measurable. The rules of this setup are: if humidity on the previous day is higher than a particular known threshold value, $x_{t-1} > X$, then it is predicted that it will rain the next day ($\hat{y}_t = 1$). The performance of this rule will then be the percentage of days for which it correctly predicts the next day's weather. This will be given by looking into the percentage of cases where the condition $x_{t-1} > X$ and $y_t = 1$ is met by the data for $t = 1, ..., T$. Call this percentage P.

In the example given, P is a statistical measure and can be formulated in an alternative way as:

$$P = Pr(y_t = 1|x_{t-1} > X)Pr(x_{t-1} > X) \tag{7}$$

The first part of the right hand side of (7) is the conditional probability of observing rain for day t given that there was a humidity higher than X on the previous day. In a practical sense, this will be given by first looking into the cases where the condition $x_{t-1} > X$ is true (call this value M), and then counting, within the M cases, the number of times in which there was rain the next day, $y_t = 1$ (call this value[6] R). Now, by calculating the fraction $\frac{R}{M}$, we have the value of $Pr(y_t = 1|x_{t-1} > X)$. But, notice that $\frac{R}{M}$ outputs the performance of the method for a restricted sample of y_t. What we really want is the percentage of correct predictions for all the T forecasted values in the data. So, if we know that the method correctly predicted R out of M cases, we can use this value to infer how well it performed for the whole sample. This is accomplished by multiplying $Pr(y_t = 1|x_{t-1} > X)$ by the percentage of cases where the condition $x_{t-1} > X$ is true. For illustration, consider the following numerical example: $T = 100$, $M = 65$, $R = 45$. These values imply that $Pr(y_t = 1|x_{t-1} > X) = \frac{R}{M} = 69\%$, $Pr(x_{t-1} > X) = 65\%$ and $P = 69\% \times 65\% = 45\%$, meaning that the forecasting rules correctly predicted 45% of the days.

Notice from the previous numerical example that a much easier way to calculate the performance of the forecasting method would be to simply divide R by T, which will also give the percentage of correct predictions (45%). It should be clear that either way we reach the same result. The difference from calculating P using equation (7) is that it represents a tractable stochastic problem, meaning that the full probability of correctly predicting y_t is a function of two other probabilities. Just as in the simple example of forecasting rain based on humidity, the tick test is a method to predict the sign of a trade (b_t) based on trade price

[6] Notice that by definition $M \geq R$.

changes (ΔP_t). Its performance is also the interaction of different conditional and unconditional probabilities, as we will show later in the paper.

In order to start the derivation, we first formalise some notation. The performance of the tick test in percentage terms is simply the number of times it got it right – that is, a correct classification of a buy (or sell), divided by the number of attempts it has made from the data. It can be represented as:

$$TT^{AC} = \frac{nCorrect^{B\&S}}{nPredictions^{B\&S}} \tag{8}$$

For equation (8) the term TT^{AC} is the accuracy of the tick test. The term $nCorrect^{B\&S}$ is simply the number of correct buy and sell classifications and $nPredictions^{B\&S}$ is the number of predictions made by the algorithm (normally equal to the number of observations).

An equivalent way of defining equation (8) is:

$$TT^{AC} = \%Correct^{B}\%Buys + \%Correct^{S}\%Sells \tag{9}$$

From a statistical point of view, we can break down the performance of the algorithm into different parts, where each will be related to a different rule of the tick test. For example, consider the first rule of the tick test – that is, if $\Delta P_t > 0$ then $\hat{b_t} = 1$. For illustrative purposes, let us suppose that we observe N trades and of these there are Z cases where we observe that the difference in trade prices for adjacent time intervals is positive. The performance of the first rule of the TT algorithm would then be, out of these Z cases, the percentage of times where there was a buy trade. If we multiplied this percentage by Z, we would get the number (and not percentage) of correct predictions for the first rule of the algorithm. We can follow the same intuition for all the other parts of TT algorithm. Notice that this is analogous to the analysis given in the example of forecasting rain.

Using this approach, we can break down the performance of the algorithm as:

$$\begin{aligned}
\%Correct^{B} \quad = \quad & Pr(b_t = 1 \mid \Delta P_t > 0)Pr(\Delta P_t > 0) + \\
& Pr(b_t = 1 \mid \Delta P_t = 0, \Delta P_{t-1} > 0)Pr(\Delta P_t = 0, \Delta P_{t-1} > 0) + \\
& \sum_{i=1}^{\infty} Pr(b_t = 1 \mid \Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} > 0) \times \\
& Pr(\Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} > 0)
\end{aligned} \tag{10}$$

and:

$$\begin{aligned}
\%Correct^{S} \quad = \quad & Pr(b_t = -1 \mid \Delta P_t < 0)Pr(\Delta P_t < 0) + \\
& Pr(b_t = -1 \mid \Delta P_t = 0, \Delta P_{t-1} < 0)Pr(\Delta P_t = 0, \Delta P_{t-1} < 0) + \\
& \sum_{i=1}^{\infty} Pr(b_t = -1 \mid \Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} < 0) \times \\
& Pr(\Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} < 0)
\end{aligned} \tag{11}$$

For equations (10) and (11), the first term on the right hand side is the probability of correctly predicting a buy or sell trade using the first part of the tick rule (cases with $\Delta P_t > 0$ and $\Delta P_t < 0$). The second term on the right hand side of (10) and (11) is related to the cases where $\Delta P_t = 0$ and $\Delta P_{t-1} > 0$. The third term in both formulas is the performance for the cases where there are continuations of zero price movements, in which the TT algorithm will seek for the last forecast made in the data. Each element of this sum will present the TT performance for each of the i^{th} zero price continuations found in the data. In general, the previous equations for the tick test performance will hold for any microstructure model and could clearly be used for extending the results in the current paper.

By following the microstructure model presented in equations (1)-(4), we can show[7] that the accuracy of the TT algorithm is a function of the spread (S), the volatility of the innovations (σ^2), the transition probabilities of trades (p) and the probability of news ($pNews$):

$$TT^{AC} = 1 + \frac{pNews\left(0.5\left(1 + erf\left(\frac{S}{\sigma\sqrt{2}}\right)(1-p)\right) - 1\right)}{1 - p(1 - pNews)} \tag{12}$$

Equation (12) explicitly defines the performance of the tick rule in percentage terms and has been confirmed by a Monte Carlo simulation. This is accomplished by defining the parameters $pNews$, p, σ^2, S and simulating T values for ϵ_t and b_t. These are used to build the process given in equations (1) to (4), which will output T artificial traded prices. The vector of price differences is set as the input for the tick test algorithm. Since the real signs of the trades are all known (see equation (4)), it is easy to compare the predicted signs of the trades from the tick test against the real ones. Needless to say, the percentage of correct predictions in this simulation exercise match the values given by equation (12) for any combination of the parameters.

In a formal analysis of equation (12), notice that the value of TT^{AC} has a lower bound of $\frac{1}{2}$ since the fraction in the right hand side of (12) is bounded by [-0.5,0]. In order to establish this result, notice that $erf\left(\frac{S}{\sigma\sqrt{2}}\right)$ takes a minimum at 0, and that the value of p that minimises TT^{AC} given that $erf\left(\frac{S}{\sigma\sqrt{2}}\right) = 0$ is $p = 1$. By plugging these values into equation (12), we can see that it takes a minimum of 0.5. This result shows that, without access to the quote data, the researcher will always do better than chance (50% of correct predictions) by using the tick test. This is an interesting result as it provides a formal argument for the use of TT algorithm in identifying the signs of trades when no quote data are available. To the best of our knowledge, this paper is the first to show such result. Notice also that the values of TT^{AC} can range from 50% to 100%, meaning that under ideal conditions, the tick test will be able to correctly predict the signs of all the trades. But it should be clear that this flawless performance will only happen in unrealistic cases (e.g. $pNews = 0$ or $pNews = 1$ and $p = 0$).

[7] See Appendix 1 for derivations.

In order to illustrate the domain of equation (12), in Figure 1 we present the shape that function TT^{AC} will take for different values of p, and S, with a fixed value of $\sigma = 1$ and $pNews = 1$.

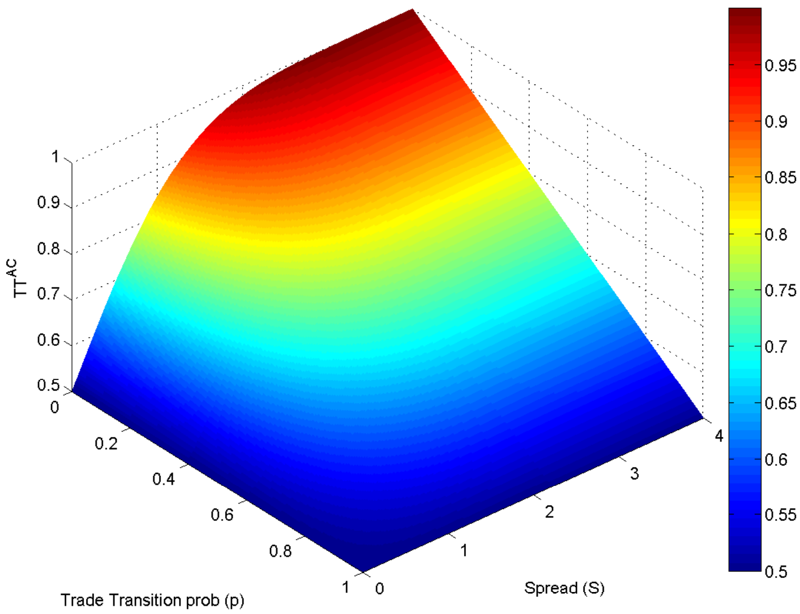

Figure 1: A 3D surface of the performance of the tick test for different values of the spread (S) and the transition probability of buy/sell (p).

The values in Figure 1 corroborate the previous analysis regarding the lower bound of the function in equation (12). As one can see, the function will reach its minimum at 0.5 and its maximum at 1. Notice also that as the spread (S) gets higher, the ratio $S/\sqrt{2\sigma^2_{\epsilon_t}}$ increases (σ is fixed at 1), resulting in a better performance of the algorithm as TT^{AC} increases. This means that the tick test will work better for assets with wide spreads and low volatility. But the effect of S and σ has a limit. Notice that after $S = 2$, the effect of this variable over TT^{AC} decreases. It should also be clear from Figure 1 that parameter p has a negative (and strong) effect on the performance of the tick test – that is, the lower the value of p, the higher the value of TT^{AC}.

3.1 An Empirical Estimation of the Performance of the Tick Test

While the microstructure model defined in (12) can capture the usual features found in tick by tick data, one of the problems on the empirical side is that the parameters from the theoretical model cannot easily be estimated from the data without further information. In this section, we outline a situation where the researcher only has access to trade data and wants to apply the equation for the tick test performance. In order to do that, we will simplify the underlying formulas with the restrictions $pNews = 1$ and $p = 0.5$. With these restrictions, the model set in (12) becomes the simplest case of a microstruture model (see Roll [1984]):

$$
\begin{aligned}
m_t &= m_{t-1} + \epsilon_t && (13)\\
P_t &= m_t + b_t \frac{S}{2} && (14)\\
\epsilon_t &\sim N(0, \sigma^2) && (15)
\end{aligned}
$$

$$
b_t = \begin{cases} 1 & \text{with 50\% probability} \\ -1 & \text{with 50\% probability} \end{cases} \tag{16}
$$

Now, assuming the last microstructure model is true, the solution for the tick test performance takes the following formula:

$$
TT^{AC} = \frac{1}{2} + \frac{1}{4} erf\left(\frac{S}{\sigma\sqrt{2}}\right) \tag{17}
$$

and empirical estimates of S and σ^2 are available using only trade data. They are defined as:[8]

$$
S = 2\sqrt{-\gamma_1} \tag{18}
$$

$$
\sigma^2 = \sigma^2_{\Delta P_t} + 2\gamma_1 \tag{19}
$$

where the parameter γ_1 is the first order autocovariance and $\sigma^2_{\Delta P_t}$ is the variance of the difference of trade prices. Both measures can be easily calculated from empirical data. With the last two results, we can directly substitute the values of S and σ^2into equation (17), which results in:

$$
TT^{AC} = \frac{1}{2} + \frac{1}{4}\left[erf\left(2\sqrt{\frac{-\gamma_1}{2\sigma^2_{\Delta P_t} + 4\gamma_1}}\right)\right] \tag{20}
$$

The robustness of such a closed formula in predicting the performance of the tick test is assessed subsequently in Section 5 of the paper.

4 The Data

The data of this study were kindly provided by the Brazilian stock exchange (Instituto Educacional BM&F Bovespa). The main database is composed of trade prices for all stocks traded on the Brazilian market from 2005 to 2011. In total, tick data for approximately 3500 assets with varying degrees of liquidity are available.

[8]See Hasbrouck [2007] and Frank and Rindi [2009] for derivations.

In the Brazilian equity market, the stocks are traded on Mega Bolsa[9] , an electronic system through which brokerage firms may execute customers' orders directly from their offices. The trading process is performed in a limit order book structure, with the usual characteristics such as price and time priority. The equity market is continuously open from 10:00 to 17:00 Brazilian time.[10] There is a break of trading between 17:00 and 17:45, and then trading re-opens for the *after market* period until 19:00.

For this study, we use the fifteen most liquids stock for the period 01/01/2009 – 01/01/2010. The degree of liquidity is measured by the number of trades. Therefore, the fifteen stocks chosen for this study are those with the highest number of trades in the selected period of time. The original data are organised in text files covering different periods (usually one month), and the following information is available in a tabular structure:

- Session Date (e.g., 2009-02-01)
- Instrument ID (ticker symbol, e.g., PETR4)
- Trade Number (e.g., 4)
- Trade Price (e.g., 24.31)
- Traded Quantity (e.g., 100)
- Trade Time (e.g., 11:30:01)
- Trade Indicator (e.g., "A", "X")
- Order Buy Date (e.g., 2005-01-02)
- Sequential Order Buy Number (e.g., 12)
- Order Sell Date (e.g., 2005-01-02)
- Sequential Order Sell Number (e.g., 14)

Most of the items in the text files are self-explanatory, with the exceptions of the trade indicator, order buy/sell date and sequential buy/sell number. The trade indicator is used only for the cases where there was a cancellation of a trade. If such a cancellation happens for different reasons (e.g., a market freeze), the trade is labelled as "A", and "X" is used for the cases where the trade is a complement of a cancellation (e.g., a cancelled buy order that generates two trades in the order book). Looking into the data, these cases are very rare.

The order buy/sell date is the date on which the trading order is generated. The Brazilian equity market has an order book structure, meaning that every trade has a corresponding buy and sell order, which was inserted by a trader/broker into the Mega Bolsa system. The date on which each buy (sell) arrived in the market is given in the item order buy (sell) date. The sequential order buy/sell number is a index that enumerates all orders each day according

[9]Recently (August 2011), the Mega Bolsa system was replaced by a new system named "Puma," which integrated all trading platforms into a single one, also increasing order processing power and decreasing trading latency.

[10]This is equivalent to UTC minus 3 (2) hours for normal (summer) time.

to the time that they occur. For example, the first order of the day (buy or sell) will always have a sequential order number of 1. This number is unique for each day, and therefore, for all types of orders, there is only one sequential number for each date. Each day, the indexing is restarted.

Note that the time of the buy and sell orders leave no doubt regarding which side is the aggressor of each trade. For example, if a buy order happens at a clock time before the sell order, then the trade is clearly a sell since it matches previously defined buy order. Therefore, the identification of buy and sell trades is done by simply observing, based on the order buy/sell date and sequential number, the order that came last. With this procedure, we can identify without any margin of error the identity of the aggressor (buyer/seller) for all trades in the data.

Dealing with high frequency data usually requires some adjustments before the statistical analysis can be pursued. This is also the case here. For the data used in the paper, all the trades which are recorded as having zero duration (no time interval between two trades) are disregarded. These zero duration trades happen when a trader buys (or sells) a large volume of the stock, consuming a relevant portion of the order book. This generates different records of trades, but the time interval between them is zero. Since this operation is relative to just one order, it should be treated as such, which justifies its deletion. We also delete the first trade of each day, since it presents a very high value of the overnight duration. Any trade with an indicator X or A, which means that they are cancelled orders, is also discarded. Table 1 presents some simple statistics regarding this adjusted sample of high frequency data.

From Table 1, we can see that the volume of data is quite high. In total there are approximately thirteen million data points. It is also possible to see that the assets are ranked according to the descending number of observations. Those with the highest number of trades are at the top of the table. This ordering is not exactly perfect given the deletion of the trades with zero duration in the treatment of the data, but it is still very clear. An almost identical ordering can also be seen for the average duration of the trades (the fourth column of Table 1). Those assets with more trades show lower values of average duration, measured as the time difference between adjacent trades.

The third column of Table 1 shows that the percentage of buyer initiated trades is around fifty percent, and so therefore is the proportion of sell trades. This sets some empirical motivation for the assumption used earlier in the theoretical part of the paper where the unconditional probability of a buy order was set equal to the unconditional probability of a sell order. The fifth column of Table 1 shows the autocorrelation coefficients for the trade signals, which are positive and statistically significant. This means that a buy (sell) trade is likely to be followed by another buy (sell) trade. This result can be explained by two different hypotheses. First, traders can split a big trading order with the objective of minimising the trade impact on the market. This action would result in a succession of trades in the same direction. The second explanation is the existence of traders with a trend-following strategy. When such traders see a large buy order executed for example, they trade in the same direction in the hope

Table 1: Descriptive statistics for the equity data used in the study

Asset	Number of Trades	Percentage of Buys Trades	Average Duration (Seconds)	Auto-correlation of Trade Signals	Auto-correlation Diff Prices
VALE5	2,007,351	48.97%	3.85	0.24	-0.28
BVMF3	1,183,625	52.88%	6.54	0.32	-0.30
GGBR4	1,053,278	50.43%	7.33	0.28	-0.22
BBDC4	942,848	51.54%	8.19	0.32	-0.23
ITSA4	927,613	52.39%	8.34	0.35	-0.32
USIM5	842,469	50.57%	9.17	0.32	-0.22
VALE3	752,938	51.42%	10.24	0.31	-0.19
ITUB4	678,629	50.38%	7.11	0.32	-0.22
PETR3	727,967	53.55%	10.53	0.32	-0.20
CSNA3	723,607	52.66%	10.68	0.37	-0.20
BBAS3	688,975	51.29%	11.22	0.30	-0.22
CYRE3	681,919	51.71%	11.33	0.32	-0.21
RDCD3	623,009	50.44%	12.34	0.33	-0.22
CMIG4	576,595	53.53%	13.33	0.34	-0.17
ALLL11	578,924	50.68%	13.35	0.32	-0.24

Notes: the first column is the asset's ticker symbol (the number stands for the type of stock and its class (e.g. ordinary, preferred and other specific classes). The second column (Number of trades) gives the number of observations in the sample. The third column gives the proportion of buyer initiated trades. The fourth column shows the average trade duration (seconds between each trade, excluding the first duration of the day). The fifth column shows the autocorrelation of the first lag of the observed/real sign of the trades. The sixth column shows the autocorrelation of the trade price differences.

that prices will continue in the same trend. It would be possible to distinguish between these explanations from the data if the unique identities of the traders were available, but this is not the case.

Another observation from Table 1 is that the autocorrelation of the price differences is negative. This result is also expected. The statistical explanation is that the traded prices implicitly contain the values of the spread (see equation (2)). This spread causes the vector of price differences to present negative autocorrelation. Formal proofs of this property can be found in the literature, e.g., Hasbrouck [2007] and Frank and Rindi [2009].

5 The Results

Next, we compare and discuss the performance of the tick test for the equity data and the performance predicted by the analytical formula given before. Table 2 shows these results.

Table 2: Empirical performance of the tick test and the performance from the analytical solution.

Asset	$\sigma_{\Delta P_t}$	γ_1	% Correct predictions (analytical)	% Correct predictions (empirical)	Absolute difference (empirical vs. analytical)
VALE5	0.0130	-0.00004688	72.08%	72.70%	0.61%
BVMF3	0.0073	-0.00001631	73.06%	79.07%	6.01%
GGBR4	0.0123	-0.00003329	69.79%	73.00%	3.20%
BBDC4	0.0155	-0.00005535	70.17%	71.22%	1.05%
ITSA4	0.0074	-0.00001747	73.60%	79.64%	6.04%
USIM5	0.0256	-0.00014052	69.52%	72.17%	2.65%
VALE3	0.0243	-0.00011055	68.13%	66.01%	2.12%
ITUB4	0.0172	-0.00006611	69.95%	68.81%	1.13%
PETR3	0.0247	-0.00012104	68.72%	64.54%	4.18%
CSNA3	0.0307	-0.00018470	68.57%	71.59%	3.02%
BBAS3	0.0152	-0.00005003	69.58%	74.40%	4.82%
CYRE3	0.0161	-0.00005401	69.21%	74.41%	5.20%
RDCD3	0.0257	-0.00014607	69.80%	72.67%	2.88%
CMIG4	0.0234	-0.00009508	67.43%	70.82%	3.39%
ALLL11	0.0117	-0.00003291	70.65%	77.35%	6.70%

Notes: the second column shows the standard deviation for the difference of traded prices (ΔP_t). The third column is the value of the first order autocovariance. The fourth column is the predicted performance of the TT algorithm calculated with the analytical formula in (20). The fifth column gives the empirical performance of the tick test. The final column shows the absolute difference of performance between the empirical test and the analytical formula.

The first feature to notice in Table 2 is that the second and third columns are the inputs used in the analytical formula of the paper – that is, $\sigma_{\Delta P_t}$ and γ_1 (see equation (20)). The empirical percentage of correct predictions by the tick test is calculated by checking all the cases where it produced a correct forecast of a buy or a sell. Remember that the true directions of the trades are easily calculated from the data by comparing the times that the buy and sell orders arrived in the market. So, by following the TT algorithm, the values in Table 2 show the percentage of times that it correctly predicted the sign of the trade for each of the stocks in the sample. In general, we can say that the empirical performance of the algorithm is positive. On average, the tick test correctly predicted approximately 72% of the trading signals.

When comparing the empirical values with those from the analytical formula, one can see that the formula presents a respectable performance. The last column of Table 2 shows the absolute difference between the analytical formula against the performance from the empirical data. The minimum value is 0.61% and the maximum is 6.7%, with an average of 3.53%. Clearly these are relatively small forecasting errors. When calculating the linear correlation of the two vectors, we get the value 0.73, which shows a very strong linear dependency between the vectors.

In order to formally assess the performance of the predictions from the analytical formula, we conduct a encompassing test. This simple methodology is based on a regression of the forecasted values against the real ones. The formula is:

$$y_t = \alpha + \beta \hat{y}_t + \epsilon_t \tag{21}$$

where y_t is the real value of the forecasted variable, in this case the empirical performance of the tick test, and $\hat{y}_t$ the respective forecasts. By using the values provided in Table 2, we use two versions of the encompassing test, one with the intercept and one without. The results are shown next:

Table 3: Results from encompassing test for the tick test performance predictions

Model	α	t-stat	β	t-stat	Wald Test Stat. ($\alpha =$ 0,$\beta = 1$)	p-value
Model 1	-0.526	-1.630	1.787	3.882	6.912	0.009
Model 2	-	-	1.037	89.156	9.984	0.007

Notes: this table shows the results from the encompassing test for the forecasting performance of the empirical analytical formula presented in the paper. Model 1 is given by $y_t = \alpha + \beta \hat{y}_t + \epsilon_t$, while model 2 can be represented as $y_t = \beta \hat{y}_t + \epsilon_t$. For both models, the dependent variable y_t is the empirical performance of the tick test (see the fifth column of Table 2 and the explanatory variable is the forecasted performance of the algorithm (see the fourth column of Table 2). The Wald test is testing the null hypothesis that α is equal to zero and β is equal to one.

, From Table 3, we can see the results from the encompassing test, which is measuring the ability of the analytical formula to predict the empirical performance of the tick test. First,the strong correlation between the forecasts and the empirical values is clearly evident. As one can see, the t-statistic of the β parameter is very high. This indicates a strong linear relationship between the forecasting model and the real measure, which, in this case, is the empirical performance of the tick test. Also, we can see that the intercept (parameter α) in model 1 is not statistically different to zero at the 10% level, meaning that the forecasting model does not possess an unconditional bias.

When looking at the results for the Wald test in Table 3, we can also see that we reject the null hypothesis in the parameter's restrictions ($\beta = 1$ and $\alpha = 0$) for both cases. If the forecasting model is unbiased, these restrictions in the parameters should hold in a statistical sense, which is not the case here. We can see from Table 3 that the values of β are higher than 1. This means that the analytical formula presented in the paper tends to underestimate the performance of the tick test. But this bias is small. The value of β in the second model is just

0.037 higher than the value of one. The small standard error of the parameter is what is driving the rejection of the null hypothesis that $\beta = 1$ in the Wald test.

While the forecasting model did not pass a formal statistical test, it should still be pointed out that the analytical formula presented relatively small values of absolute errors when compared to the real performance (see Table 2) and that it presented a very strong linear correlation with the measure of interest (see Table 3). Given our objective to forecast the performance of the tick test algorithm without any access to the real trade signs, we believe that the performance of the method in the research is still quite impressive, although statistically it incurred a small forecasting bias.

6 Conclusions

In this paper we set out to investigate the performance of the tick test from an analytical point of view. Based on a general microstructure model, we derived analytical formulae for the performance of the algorithm and we also made available a simpler formula that can be easily applied to empirical data, even in the absence of quote prices. The main formula derived indicates the existence of boundaries on the performance of the tick test, and specifically, it is shown that it will always perform better than chance (50% correct predictions). The accuracy of a reduced version of the analytical formula was tested using fifteen highly liquid stocks from the Brazilian equity market, which demonstrated that the derived equation well represents the empirical performance one would get when using the tick test.

The implications of this research are clear. Studies including those of Boehmer et al. [2006], and Tanggaard [2003] have showed that by using trade inference algorithms in the construction of microstructure variables such as the PIN, there will be an implied bias. The extent of this bias is a direct function of the ability of the trade sign prediction algorithm to make correct forecasts. Our paper provides a way of assessing such performance based on a simple formula.

A further advance of the present research would be devise a way to estimate the parameters of the full microstructure model based on incomplete data, which would allow for the empirical application of the original, full model for examination of the tick test performance.

Appendix 1 - Analytical Derivation of the Performance of the Tick Test

The accuracy of the algorithm, in percentage terms, is given by:

$$TT^{AC} = \%Correct^{B}\%Buys + \%Correct^{S}\%Sells \tag{22}$$

where:

$$\%Correct^{B} = A^{B} + B^{B} + C^{B} \tag{23}$$

$$\begin{aligned}
A^{B} &= Pr(b_t = 1 \mid \Delta P_t > 0)Pr(\Delta P_t > 0) && (24)\\
B^{B} &= Pr(b_t = 1 \mid \Delta P_t = 0, \Delta P_{t-1} > 0)Pr(\Delta P_t = 0, \Delta P_{t-1} > 0) && (25)\\
C^{B} &= \sum_{i=1}^{\infty} Pr(b_t = 1 \mid \Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} > 0) \times \\
&\quad Pr(\Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} > 0) && (26)
\end{aligned}$$

and:

$$\%Correct^{S} = A^{S} + B^{S} + C^{S} \tag{27}$$

$$\begin{aligned}
A^{S} &= Pr(b_t = -1 \mid \Delta P_t < 0)Pr(\Delta P_t < 0) && (28)\\
B^{S} &= Pr(b_t = -1 \mid \Delta P_t = 0, \Delta P_{t-1} < 0)Pr(\Delta P_t = 0, \Delta P_{t-1} < 0) && (29)\\
C^{S} &= \sum_{i=1}^{\infty} Pr(b_t = -1 \mid \Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} < 0) \times \\
&\quad Pr(\Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} < 0) && (30)
\end{aligned}$$

For equation (22), the value of $\%Buys$ is simply the unconditional probability of a buy order:

$$\%Buys = Pr(b_t = 1) = Pr(ST_t = 1) = \frac{1}{1 - \frac{1-p_{Buy}}{p_{Sell}-1}} \tag{31}$$

By equating the unconditional probability of a buy order to the unconditional probability of a sell order ($Pr(b_t = 1) = Pr(b_t = -1) = 0.5$), we get the result that the transition probabilities of a buy order is equal to the transition probability of a sell order, that is $p_{Buy} = p_{Sell} = p$.

Now, starting the derivations of the first conditional probability in (23) we can show using Bayes rule that in the formula for A^{B} the first term of the equation will take the following shape:

$$Pr(b_t = 1 \mid \Delta P_t > 0) = \frac{Pr(\Delta P_t > 0 \mid b_t = 1)Pr(b_t = 1)}{Pr(\Delta P > 0)} \tag{32}$$

This modification simplifies the derivations as working by conditioning on the event $b_t = 1$ is easier than working by conditioning on the event $\Delta P_t > 0$. Expanding the conditional probability given before by following the underlying microstructure model we can show that:

$$\begin{aligned} Pr(b_t = 1 \mid \Delta P_t > 0) &= pNews\,(0.5p + (1-p)(1-\Phi(-S))) + \\ &\quad (1-pNews)(1-p) \end{aligned} \tag{33}$$

where $\Phi(x)$ is the Normal cumulative distribution with variance σ^2 evaluated at x. Following last equation and remembering that $Prob(b_t = 1) = 0.5$, we have the first part of equation (23) worked out. For the second term in (23), B^B, we can show by using Bayes formula that:

$$Pr(b_t = 1 \mid \Delta P_t = 0, \Delta P_{t-1} > 0) = \frac{Pr(\Delta P_t = 0, \Delta P_{t-1} > 0 \mid b_t = 1) Pr(b_t = 1)}{Pr(\Delta P_t = 0, \Delta P_{t-1} > 0)} \tag{34}$$

By deriving the conditional probability with respect to the underlying microstructure model, we have the following result:

$$\begin{aligned} Pr(\Delta P_t = 0, \Delta P_{t-1} > 0 \mid b_t = 1) &= [pNews\,(0.5p + (1-p)(1-\Phi(-S))) + \\ &\quad (1-pNews)(1-p)]\,(1-pNews)(p) \end{aligned} \tag{35}$$

Now, notice that equation (35) is equivalent to equation (33), except for the multiplying term $(1-pNews)(p)$. This is the effect of the extra condition on the conditional probabilities of a buy order. This property also holds when working with the other conditions in the third element on the right hand side of equation (23). By working out the elements of the sum in C^B, a pattern becomes clear. Formalising such a pattern gives:

$$Pr(\Delta P_t = 0, .., \Delta P_{t-i} = 0, \Delta P_{t-i-1} > 0 \mid b_t = 1) = \quad A^B\,[(1-pNews)(p)]^i \tag{36}$$

It is clear from equation (36) that the performance of the second rule of the tick test is related to the performance of the first rule (a value of Z). The value of A^B is adjusted by the parameter i.

It should also be pointed out that one property used in the paper is that:

$$\sum_{i=0}^{\infty} [(1-pNews)(p)]^i = \frac{1}{1-(1-pNews)(p)} \tag{37}$$

which greatly simplifies the analysis by removing the sum part in the derivations. The derivations for the other part of the tick test algorithm, that is, equation (23) are very similar to the derivation of the elements of equation (22) and are therefore not given here.

By placing all the previous results back in the original formula and simplifying, we have the final equation for the tick test performance:

$$TT^{AC} = 1 + \frac{pNews\left(0.5\left(1 + erf\left(\frac{S}{\sigma\sqrt{2}}\right)(1-p)\right) - 1\right)}{1 - p(1 - pNews)} \tag{38}$$

References

M. Aitken and A. Frino. The accuracy of the tick test: Evidence from the australian stock exchange. *Journal of Banking & Finance*, 20(10):1715–1729, 1996.

P. Asquith, R. Oman, and C. Safaya. Short Sales and Trade Classification Algorithms. *SSRN eLibrary*, 2008.

E. Boehmer, J. Grammig, and E. Theissen. Estimating the Probability of Informed Trading - Does Trade Misclassification Matter? *SSRN eLibrary*, 2006.

S. Chakravarty. Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics*, 61(2):289–307, 2001.

T. Chordia, R. Roll, and A. Subrahmanyam. Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, 76(2):271–292, 2005.

T. Chordia, R. Roll, and A. Subrahmanyam. Liquidity and market efficiency. *Journal of Financial Economics*, 87(2):249 – 268, 2008.

D. Easley, N. M. Kiefer, M. O'Hara, and J. B. Paperman. Liquidity, information, and infrequently traded stocks. *The Journal of Finance*, 51(4):1405–1436, 1996.

K. Ellis, R. Michaely, and M. O'Hara. The accuracy of trade classification rules: Evidence from the nasdaq. *The Journal of Financial and Quantitative Analysis*, 35(4):529–551, 2000.

T. J. Finucane. A direct test of methods for inferring trade direction from intraday data. *Journal of Financial and Quantitative Analysis*, 35(04):553–576, December 2000.

D. J. Frank and B. Rindi. *The Microstructure of Financial Markets*. Cambridge University Press, Cambridge, UK, 2009.

J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207, 1991.

J. Hasbrouck. *Empirical Market Microstructure*. Oxford University Press. New York, 2007.

C. M. C. Lee. Market integration and price execution for nyse-listed securities. *Journal of Finance*, 48(3):1009–38, 1993.

C. M. C. Lee and M. J. Ready. Inferring trade direction from intraday data. *Journal of Finance*, 46(2):733–46, 1991.

E. Odders-White. On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets*, 3:259–286, 2000.

M. Peterson and E. Sirri. Evaluation of the biases in execution cost estimation using trade and quote data. *Journal of Financial Markets*, 6:259–280, 2003.

R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4):1127–39, 1984.

C. Tanggaard. Errors in trade classification: Consequences and remedies. Finance Working Papers 03-6, University of Aarhus, Aarhus School of Business, Department of Business Studies, 2003.

E. Theissen. A test of the accuracy of the lee/ready trade classification algorithm. *Journal of International Financial Markets, Institutions and Money*, 11:1416–5, 2000.