# Estimating the Intensity of News Based on Trade Data

Marcelo Perlin[1] (marcelo.perlin@ufrgs.br)

**Subject:** Risk management and volatility (Gestão de risco e volatilidade)

[1] Assistant Professor, Federal University of Rio Grande do Sul (Brazil). Acknowledgements: The author is grateful for the helpful comments received from the participants of the microstructure sessions in the 2013 EFMA Conference in Reading UK. The author also acknowledges and is very grateful for the financial support from Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS), process number 0854 12-3 (ARD-003/2012)}

# Estimating the Intensity of News Based on Trade Data

**Abstract:** This paper investigates the problem of identifying the strength of incoming news in the financial market. Based on a microstructure model we are able to derive a simple formula that estimates the likelihood of having news for any given tradable asset in a particular time period. In the empirical part of the paper, we investigate the properties of this proposed estimator for the Brazilian equity market. We find that the strength of news has a common (systematic) component across all assets, a strong and positive correlation to volatility of price differences and a clear intraday pattern across all assets.

**Keywords:** market microstructure, effect of news, volatility

# Introduction

The role of news is fundamentally important in the financial literature as it is one of the main factors that moves prices. We can define news as any novel information (true or not) becoming public and changing investor's expectations about the future cash flows of financial instruments. This includes, but is not restricted to the publication of financial statements, the definition of a new interest rate by the government, the disclosure of a new product by a company or even the death of a company's CEO. The relationship of these events to financial prices is that they change the perception of future value by the investors, who in turn update their quotes and change the observed market prices of financial contracts. If most of the investors believe that a recent event will increase/decrease future cash flows, prices are likely to go up/down.

The subjective way in which investors perceive a particular event is related to a well-know measure in finance and economics, price volatility. As an example, the death of a CEO can be seen as good news by some and bad news by others. The asymmetry of perception between market participants induces higher price uncertainty as traders invest based on their expectations. This means that prices can move faster within a particular time interval when there is new information in the market. Volatility is, therefore, a by-product of news. Not surprisingly, the empirical effects of news (and volatility) in the financial market covers a significant proportion of the academic literature in finance.

In the early days, we can find the work of Fama et al (1969) which studies the resulting impact of stock splits in trade prices, finding that stock splits do bring new information to the market since it is usually associated with higher dividends. Waud (1970) follows the same type of investigation as the previous authors, but studies the effect of news with respect to the American discount rate. The author also finds a significant impact of changes in the discount rate towards stock market returns.

As we can expect, the last two papers have been followed by a significant amount of research on the case study of the impact of news. We emphasise the work of Pearce and Roley (1985), which studies the impact of surprise announcements based on a data set of investor survey data and news in the media regarding several economic indicators such as money supply, inflation and the discount rate. The authors find strong evidence that unexpected news (surprises) for monetary policy affects the stock market. In more recent years, we can find several studies on this topic. Just to cite a few, we have Mitchell and Mulherin (1994) on the relationship of the amount of news announcements and market activity, Barber and Odean (2008) on the effect of media attention over trade prices, Carvalho, Klagge and Moench (2011) on the case study of the persistence effect of false news, Birz and Lott Jr. (2011) on the effect of newspapers articles over stock returns, among many other research articles[2]. These studies provide a consistent body of empirical evidence regarding the effect of news in the financial market under different perspectives.

In general, the previous literature on news impact can be summarised in two steps: the incoming news is identified in the different media (e.g., newspaper, internet) and its impact is measured over different financial variables, either by using an event study or the estimation of a particular econometric model. In the present paper, we take an alternative route. We approach the study of the impact of news on prices as a problem of incomplete data. The change of trade prices over time is related to two components, news (efficient price) related changes and market frictions (or microstructure noise). We identify these two components in the price formation of

[2] See Wasserfallen (1989), Becker, Finnerty, and Friedman (1995), Boyd, Jagannathan and Hu (2001), Rangel (2011), Engelberg and Parsons (2011), Garcia (2013), among others.

a microstructure model and set out to estimate the intensity of the incoming news based only on the available data. That is, we observe the prices of trades executed in a particular time period and set out to identify the strength of news in the market based solely on that information. This is an interesting approach as we treat news as generic information coming to the market and updating investors' beliefs about the fair price of the asset. From the scientific point of view, this is a more objective approach than retrieving news from the media since there is no assumption of what is defined as news as we let the data speak for itself.

We point out that this is a novel approach in the literature of news impact. The proposed method does not rely on the gathering of media data such as newspaper articles or any other channel. It is a data based method and all that is necessary to implement it is level 1 data, trade prices and trade signs (identities of trade aggressors [buyers or sellers]). As will be shown in the rest of the paper, the proposed indicator of news is surprisingly easy to calculate, taking only one input.

In the empirical part of the paper, we investigate the properties of this estimator on a daily basis for 20 stocks of the Brazilian equity market over an extensive period of two years. We find that intensity of news has a strong common component across all stocks. We attribute this to the incoming of news regarding the Brazilian economy, which will in turn systematically affect all stocks in the sample. From our second regression model, we report that news intensity has a tendency to cluster, with a positive relationship to the volatility of price changes and a negative correlation to trade volume. We attribute the first two results to the fact that volatility is a bi-product of news and, therefore, both will present a positive correlation, with a clustering pattern. We explain the negative association of news intensity to trade volume as the effect of informed traders hiding their whole order by trading smaller volumes.

Another interesting result from the study is the existence of an intraday profile of news intensity. By aggregating the estimator of news intensity based on the time of day, we find that in the beginning of trading hours there is a relatively higher value of incoming news. This is an intuitive result as the beginning of trading hours is exactly the time of day when there is a higher volume of new information to be incorporated into prices.

In general, the results from this research show that the proposed estimator of news intensity has the expected properties when confronted with real data, showing its potential for financial research. In summation, this paper contributes to the literature by first proposing a method for quantifying the strength of news in the market, and second, by showing that it has the statistical properties one would expect for this particular type of variable.

## Estimating the intensity of news

The idea of a microstructure model is to formalize the way in which a price comes to the market. In the paper, we suggest an improved version of Roll's microstructure model (De Jong & Rindi, 2009; Hasbrouck, 2007). The novel model will be given by the following set of equations:

$$m_t = m_{t-1} + \epsilon_t \qquad (1)$$

$$P_t = m_t + b_t \frac{S}{2} \qquad (2)$$

$$\epsilon_t \sim \begin{cases} = 0 \text{ with probability } 1 - pNews \\ \neq 0 \text{ with probability p}News \end{cases} \tag{3}$$

$$b_t \sim \begin{cases} = 1 \text{ if } \mathrm{ST}_t = 1 \\ -1 \text{ if } \mathrm{ST}_t = 2 \end{cases} \tag{4}$$

with $ST_t$ following a Markov chain with transition probabilities given by matrix $P$ :

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \tag{5}$$

For Equation (1), the term $m_t$ is the efficient (true) price of an asset, which will follow a random walk representation with innovations given by $\epsilon_t$. Different from the previous case, variable $\epsilon_t$ will follow a discrete probability distribution, with zero expectation. Notice that the underlying model has a superficial description of the distribution of the innovations, with no assumption whatsoever about the shape of it. In fact, the assumption of the innovation's distribution is not necessary for the derivations in the paper. The only requirement was that it must be a discrete distribution with zero expectation. Unlike in the continuous case, the discreteness assumption will allow for the innovation to take any value $(Pr(\epsilon_t = x) \neq 0)$. This is a more realistic assumption as the trade prices are usually quoted in decimal units in the financial markets.

Now, an important innovation in the proposed model is that it allows for the explicit parametrisation of the intensity of news in the market $pNews$, which can take any value between $0$ and $1$. Sometimes there will be news in the market ($\epsilon_t \neq 0$) and sometimes not ($\epsilon_t = 0$). The higher the value of $pNews$, more intense is the incoming news, meaning that one can expect a higher volume of new information reaching the market per unit of time.

The second equation in the system, $P_t$, represents the actual price of a trade. The change here is that the trade direction will follow a Markov chain which can mimic the empirical autocorrelation in trade signs. The explicit representation of the trade signs as a Markov chain is not novel (Choi, Salandro, & Shastri, 1988). The use of the same transition probability $p$ for buys and sells is based on empirical observations of trade data, which shows that the proportion of buys and sells is well balanced, meaning that unconditionally, a buy trade is equally likely as a sell.[3] We point out that the simpler model of (Roll, 1984) is nested with a

[3] From the mathematical point of view, 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 $p_1 = p_2$.}

discrete (and unknown) distribution for $\epsilon_t$ in the proposed microstructure model by setting $pNews = 1$ and $p = 0.5$.

Now, the main idea of this paper is to estimate the value of $pNews$ based only on trade data ( $P_t$ ). The reader can find the full derivation exercise in Appendix A. Our final result is:

$$pNews = 1 - Pr\left(\Delta P_t = 0 \mid \Delta b_t = 0\right) \quad (6)$$

That is, the probability of news is related to the probability of a zero price movement, conditional on the cases where the trade signs between $t$ and $t-1$ are equal. This conditional probability can be easily calculated for empirical data by first restricting the cases of zero trade price differences only when the trade signs are equal in one period to the other. The simplicity of this estimator of news intensity is one of its selling points. In the next part of the paper, we investigate the properties of estimated values of $pNews$ for the Brazilian equity market.

# The Data

The data of this study is composed of trade prices of stocks from the Brazilian stock exchange (Bovespa) and it was kindly provided by Instituto Educacional BM&F Bovespa. This is a very dense database with tick data for approximately 3,500 financial instruments, including stocks and some derivatives from the year 2005 to 2012. In the Brazilian equity market, the stocks are traded 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.[4] A break of trading occurs 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 twenty stocks with the highest number of trades in the time period between year 2010 and 2012. These are the most liquid assets, which justifies our criteria. This selection implies a trading frequency of approximately 4.075 trades per stock, per day.

High frequency trading data usually has some unwanted properties that should be removed prior to the statistical analysis. For the data used in the study, all the trades with zero duration (no time interval between two trades) are removed. This happens when a large market order reaches the order book and consumes a large portion of it. However, even though this is a single trade, the data represents it as different events (trades) with the same time interval (and, therefore, zero duration). These cases are removed from the sample. We also delete any cancelled trade and the first trade of each day since it has a high duration. Table 1 shows some simple statistics for this data after the pre-processing.

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

Table 1 - Descriptive Statistics of the Data

| Asset | Average Number of Observations per day | Average Duration | Autocorrelation Diff Price (1st lag) | Autocorrelation Trade Signals (1st lag) |
|---|---|---|---|---|
| VALE5 | 8045,687 | 4,130 | -0,290 | 0,247 |
| OGXP3 | 6148,384 | 5,754 | -0,308 | 0,261 |
| BVMF3 | 5725,783 | 6,475 | -0,341 | 0,293 |
| ITUB4 | 5163,679 | 6,793 | -0,265 | 0,272 |
| PDGR3 | 4938,179 | 8,002 | -0,307 | 0,295 |
| GGBR4 | 4276,052 | 8,132 | -0,273 | 0,273 |
| BBAS3 | 4350,452 | 8,273 | -0,248 | 0,271 |
| BBDC4 | 4249,775 | 8,257 | -0,255 | 0,262 |
| ITSA4 | 4336,817 | 8,430 | -0,336 | 0,300 |
| USIM5 | 3899,576 | 9,037 | -0,272 | 0,269 |
| PETR3 | 3366,215 | 10,796 | -0,246 | 0,298 |
| CYRE3 | 3679,153 | 9,534 | -0,234 | 0,290 |
| GFSA3 | 3371,978 | 11,149 | -0,302 | 0,285 |
| VALE3 | 3224,996 | 10,760 | -0,222 | 0,288 |
| MRVE3 | 3295,837 | 10,925 | -0,246 | 0,292 |
| CSNA3 | 2841,291 | 12,344 | -0,234 | 0,293 |
| RSID3 | 2691,827 | 13,604 | -0,256 | 0,315 |
| HYPE3 | 2582,149 | 17,385 | -0,245 | 0,288 |
| CIEL3 | 2689,612 | 13,597 | -0,261 | 0,285 |
| RDCD3 | 2640,847 | 14,254 | -0,238 | 0,277 |

Notes: the first column is the asset's ticker symbol where the number stands for the type of stock and its class (e.g., ordinary, preferred and other specific classes). The second column shows the average number of trades for each day in the sample. The third column gives the average trade duration (seconds between each trade, excluding the first duration of the day). The fourth column shows the autocorrelation of the first lag of trade prices differences. The fifth column shows the autocorrelation of the observed/real signs of the trades (+1 for a buy, -1 for a sell).

From Table 1, we can see the high density of the data, which is a particular feature of high frequency data. In total, covering all stocks, there are approximately 50 million data points. The order of appearance of the assets in Table 1 respects our liquidity measure, the number of trades. Stocks with the highest/lowest number of trades are at the top/bottom of the table. The exclusion of zero duration trades affected this ordering, but the property is still very clear. The average duration (third column) also respects this ordering pattern as the assets with the most trades are also the ones with the lowest values of average duration.

We can see from Table 1 that the autocorrelation for the price differences (fourth column) is negative for all stocks. This means that a trade price increase/decrease is most likely to be followed by an inversion, a decrease/increase. This is a well-know property of trade prices which implicitly contains the values of the spread. This causes the negative autocorrelation for the price differences. More details about this point can be found in the main literature, De Jong and Rindi (2009) and Hasbrouck (2007).

The sixth column of Table 1 shows the autocorrelation coefficients for the trade signals, which are all positive. This means that a buy/sell trade is likely to be followed by another buy/sell trade. The result of a positive autocorrelation for trade signals corroborates with the arguments presented before for justifying the use of a Markov chain as the stochastic process of trades.

# Methodology

In this section, we are interested in studying the properties of the proposed measure of news intensity developed in the first part of the paper. From Equation (6), we can see that it is a measure of the likelihood of news for a particular time period. In order to give it some dynamic, we calculate the likelihood of news within the periodicity of a day. That is, for each day we calculate the conditional probability of a zero trade price difference and use it as input it at Equation (6). From the statistical point of view, this is given by:

$$pNews_i = 1 - Pr(\Delta P_{i,t} = 0 \mid b_{i,t} = b_{i,t-1}) \qquad (7)$$

For last equation, the term $pNews_i$ is the probability of news for a particular stock in day $i$, where $i$ goes from $1$ to 498 (number of trading days between the years of 2010 and 2012).

The empirical analysis of the paper is divided into two sections. In the first part, we seek out to understand the existence of commonalities in the probability of news for the different assets. Since all the stocks in the dataset are from the Brazilian market, we can expect that a portion of the news in each of them is related to the incoming news regarding the Brazilian market as a whole. This can be either the disclosure of economic reports, a new presidential election, and so on. If this is correct, then there must be a common factor in the probability of news of all assets over time. We test this hypothesis with the estimation of the following model:

$$pNews_{i,j} = \alpha_j + \phi_j pNews_{i,j}^{Agg} + \epsilon_t \qquad (8)$$

For Equation (8), the term $pNews_{i,j}^{Agg}$ is the average of the probability of news across the assets over time. To explain this more clearly, $pNews_{i,j}$ is a matrix with 498 rows and 20 columns. We calculate $pNews_{i,j}^{Agg}$ by excluding column $j$ and averaging the likelihood of news over the rows. This provides an approximation for the probability of news for the market as a whole. If our hypothesis of a common movement is correct, then parameter $\phi_j$ should be positive and statistically significant.

In the second part of the empirical section, we try to explain the behaviour of the news intensity indicator with respect to other variables, such as volatility, trading volume and others. Formally, this is accomplished with the estimation of the following econometric model:[5]

$$pNews_{i,j} = \alpha_j + \beta_{j,1} pNews_{i-1,j} + \beta_{j,2} Volat_{i,j} + \beta_{j,3} Vol_{i,j} + \beta_{j,4} dur_{i,j} + \epsilon_{i,j} \qquad (9)$$

The basic idea in the estimation of Equation (9) is to investigate how the estimated likelihood of news relates to other variables and itself. As we argued in the introduction of the paper,

[5] Given the limited range of the dependent variable in Equation (9), a non-linear (logit) version of the econometric model was also estimated. The results with respect to the signs and statistical significance of the parameters are quite similar to the linear version.

volatility is bi-product of news. We also know that volatility has a tendency to cluster.[6] Likewise, we should also expect such a property for the intensity of news since it is one of the main components of volatility. Following this argument, we add an AR(1) term in order to check for the clustering of news intensity. This property was already illustrated in Figure 1, but we include the autoregressive parameter in order to test it formally for our data.

We also add contemporaneous volatility, trade volume and duration in our econometric analysis. For Equation (9) the independent variables, with exception of $pNews_{i-1,j}$, are all averages taken across the day. The volatility ($Volat_{i,j}$) is the standard deviation for the log returns of traded prices in day $i$. The term $Vol_{i,j}$ is the average volume of trades divided by 10.000 for day $i$ and $dur_{i,j}$ is the average duration (time between trades), divided by 100.

# The Results

We start the presentation for the first model, Equation (8).

Table 2 - Results from Econometric model, Equation (8)

| Asset | $\alpha_j$ | $\phi_j$ | Adj R2 |
|---|---|---|---|
| VALE5 | 0,15*** | 0,63*** | 0,25 |
| OGXP3 | 0,20*** | 0,39*** | 0,04 |
| BVMF3 | 0,03 | 0,68*** | 0,08 |
| ITUB4 | 0,08** | 0,84*** | 0,21 |
| PDGR3 | -0,12*** | 1,13*** | 0,17 |
| GGBR4 | -0,10*** | 1,26*** | 0,32 |
| BBAS3 | 0,16*** | 0,60*** | 0,09 |
| BBDC4 | 0,12*** | 0,76*** | 0,16 |
| ITSA4 | 0,00 | 0,76*** | 0,12 |
| USIM5 | -0,10*** | 1,36*** | 0,26 |
| PETR3 | 0,08* | 0,84*** | 0,15 |
| CYRE3 | 0,07* | 0,84*** | 0,14 |
| GFSA3 | -0,03 | 0,93*** | 0,14 |
| VALE3 | 0,14*** | 0,88*** | 0,20 |
| MRVE3 | 0,14*** | 0,60*** | 0,07 |
| CSNA3 | 0,10** | 0,84*** | 0,16 |
| RSID3 | 0,11*** | 0,66*** | 0,08 |
| HYPE3 | -0,06 | 1,23*** | 0,14 |
| CIEL3 | 0,50*** | -0,32* | 0,01 |
| RDCD3 | 0,21*** | 0,57*** | 0,05 |

Notes: The econometric model is given by $pNews_i = \alpha_j + \phi_j pNews_i^C + \epsilon_t$, where $pNews_i^C$ is the average of $pNews_i$ across all assets, excluding asset $i$. All the standard errors used for statistical testing are robust to

[6] See Arch/Garch models, Engle (1982) and Bollerslev (1986).

heteroskedasticity and serial correlation (Newey & West, 1987). The values with *, ** and *** means statistical significance at 10%, 5% and 1%, respectively.

The results from Table 2 show a positive relationship between the news intensity of the individual assets and the news intensity of the market as a whole. For all assets except one, parameter $\phi_j$ is positive and statistically significant at 1%. This strongly suggests the existence of a common factor for the likelihood of news for all assets. Again, this was an expected and intuitive result since all the stocks have a country-related risk factor in common.

The second result we have is for the investigation of the relationship of the news intensity variable with respect to other commonly used variables in financial research.

Table 3 - Results from Econometric model, Equation (9)

| Asset | $\alpha_j$ | $\beta_{j,1}$ | $\beta_{j,2}$ | $\beta_{j,3}$ | $\beta_{j,4}$ | Adj R2 |
|---|---|---|---|---|---|---|
| VALE5 | 0.21*** | 0.14*** | 9.12*** | -0.15 | 0.95 | 0.57 |
| OGXP3 | 0.11*** | 0.15*** | 4.58*** | 0.05 | 7.17*** | 0.41 |
| BVMF3 | 0.05 | 0.11*** | 7.14*** | 0.09*** | 4.04*** | 0.68 |
| ITUB4 | 0.21*** | 0.12*** | 7.30*** | -0.40*** | 1.33*** | 0.47 |
| PDGR3 | 0.08*** | 0.19*** | 3.31*** | -0.26*** | 6.45*** | 0.47 |
| GGBR4 | 0.05 | 0.43*** | 4.49*** | 0.47*** | 5.34*** | 0.48 |
| BBAS3 | 0.22*** | 0.07*** | 6.33*** | -0.25 | -0.15 | 0.45 |
| BBDC4 | 0.20*** | 0.15*** | 6.82*** | -0.35*** | -0.25 | 0.50 |
| ITSA4 | 0.02*** | 0.05 | 7.17*** | -0.14*** | 3.99*** | 0.75 |
| USIM5 | 0.14*** | 0.49*** | 1.82*** | -0.82*** | 5.49*** | 0.39 |
| PETR3 | 0.17*** | 0.09*** | 7.46*** | 0.15 | -0.25 | 0.56 |
| CYRE3 | 0.13*** | 0.14*** | 4.75*** | -0.16*** | 4.00*** | 0.55 |
| GFSA3 | 0.09*** | 0.33*** | 2.13*** | -0.42*** | 4.49*** | 0.38 |
| VALE3 | 0.29*** | 0.14*** | 6.37*** | -0.57*** | -0.45 | 0.41 |
| MRVE3 | 0.18*** | 0.11*** | 3.28*** | -0.44*** | 1.45 | 0.43 |
| CSNA3 | 0.14*** | 0.33*** | 2.64*** | 0.35 | 2.54*** | 0.30 |
| RSID3 | 0.14*** | 0.08*** | 3.34*** | 0.05 | 2.32*** | 0.53 |
| HYPE3 | 0.20*** | 0.21*** | 2.04*** | 0.05 | 0.75 | 0.36 |
| CIEL3 | 0.18*** | 0.44*** | 1.81*** | -0.53*** | 0.35 | 0.38 |
| RDCD3 | 0.24*** | 0.09*** | 3.95*** | -0.25 | 0.05 | 0.44 |

Notes: The econometric model is given by $pNews_{i,j} = \alpha_j + \beta_{j,1} pNews_{i-1,j} + \beta_{j,2} Volat_{i,j} + \beta_{j,3} Vol_{i,j} + \beta_{j,4} dur_{i,j} + \epsilon_{i,j}$, where $Volat_{i,j}$ is the average volatility (standard deviation) of price changes for day $i$, $Vol_{i,j}$ is the average volume of trades and $dur_{i,j}$ is the average duration (time between trades. All the standard errors used for statistical testing are robust to heteroskedasticity and serial correlation (Newey & West, 1987). The values with *, ** and *** means statistical significance at 10%, 5% and 1%, respectively.

The first result we have from Table 3 is for the autoregressive property of news intensity. The value of $\beta_{1,j}$ is positive and significant for 95% of the cases (19 out of 20). This means that a high probability of news in a day is most likely to be followed by another high probability of news. This corroborates with the visual inspection of Figure 1, where it was possible to visualize the clustering property of news intensity. The contemporaneous correlation of volatility and news can also be seen in Table 3, where parameter $\beta_{j,2}$ is positive and significant for all cases of the data. This result provides statistical evidence for the positive relationship between news and the volatility of price changes.

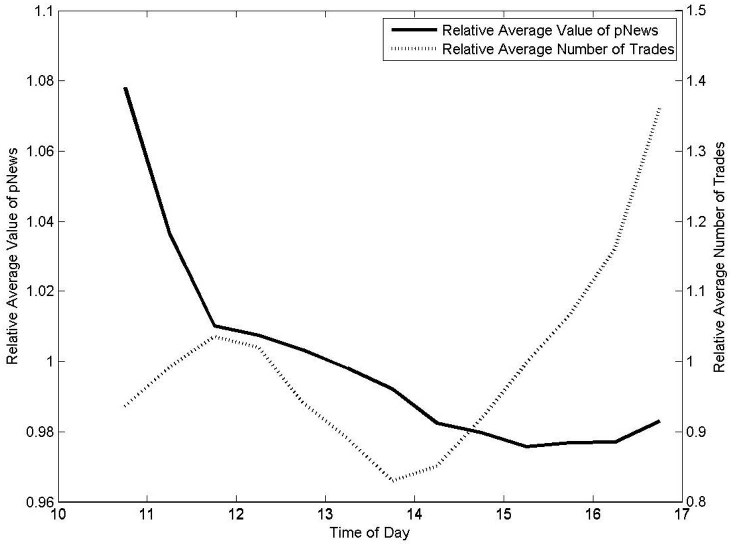
An interesting result is found for the relationship between average volume of trades and the probability of news. We see from Table 3 that the values of $\beta_{j,3}$ are mostly negative. Out of 20 cases,10 are negative and statistically significant. While this result is not particularly strong, it still implies that a high likelihood of news is related to a low volume per trade. This is an interesting result and can be explained by microstructure theories. The presence of news in the market is also related to the presence of informed traders, which have the right motivation to avoid the disclosure of their private information. They can do so by trading small quantities of the asset; that is, they fragment their whole order. This effect will then imply a negative correlation between trade volume and intensity of news, which is the result we find.

When looking at parameter $\beta_{j,4}$, we find that the relationship between duration and news intensity is mostly positive, where 11 cases presented statistically significant coefficients with positive values. This is an interesting, but not intuitive result as one can expect that a higher intensity of news is related to a low duration and higher trading activity. However, this unexpected result can be explained by the intraday profile of news, which is the next analysis taken in the paper.

In this study, we also look at the intraday profile of the news intensity estimator. This particular investigation is motived by the fact that the market has a particular structure of incoming of news due to their opening and closing hours. For evidence of this effect, in the paper of Grob-Klubmann and Hautsch (2011), such a profile was presented by comparing the volume of news with respect to the time of day it was announced. Therefore, one should expect that this intraday profile will also affect the values of the estimated intensity of news.

We conduct our analysis of the intraday pattern of $pNews$ based on all 20 assets using a simple aggregation procedure. First we divide the whole trading day into 30-minute intervals, calculating the number of trades and intensity of news with Equation (7), and then averaging it for each time interval. This gives an intraday profile of $pNews$ and number of trades for each stock. The second step is to divide each stock's value of intraday $pNews$ and number of trades by its own sample average, which then gives percentage deviations from the mean, across each time interval. After that, we take another average of the percentages for the thirty minutes periods across all assets, resulting in the next figure.

Figure 2 - Aggregated relative values of intraday *pNews* for all 20 assets in the database




From Figure 3 we can see that the aggregated dynamic for intraday relative values of *pNews* follows a decaying dynamic, with the news intensity at the beginning of the day with at least 8% higher value of relative the sample's mean[7]. In Figures 2 we also add the number of trades (right axis) as a proxy for trading intensity. The reported "U" shape of the trading intensity was expected as this is a standard result in empirical market microstructure.[8] Now, it should be pointed out that an increase of trading intensity is an indication of the presence of informed trades and, thereby, the incoming of news (Dufour & Engle, 2000), where traders respond to the new set of available information, adjust their portfolios according to their expectations and increase trading intensity.

While the values in Figure 2 show a correlation between higher trading intensity and higher news intensity at the beginning of the day, it does not present the same result at the end of the day, where news intensity decreases while the trading intensity increases. This result is not particularly intuitive and we can present an alternative explanation by justifying it as the effect of the inventory management of overnight risk and trading costs.

It is well-know that traders do not like to bear overnight risks in their positions. Therefore, they are likely to close their trades by the end of the trading day. Notice that this is not related to incoming new information, but due to the avoidance of overnight market risk. Also, particular to the case of Brazil, regulation by the exchange sets that the trading costs are lower if a trade in the stock market is opened and closed in the same day.[9]. This means that traders in Brazil have not only a risk motive for closing the trades at the end of the day, but also a financial motive. Following this logic we can argue that most trades at the end of the day are not news-oriented but simply motivated by the avoidance of risk and the minimization of costs. This can

[7] One can see the same intraday pattern for real (and not estimated) news announcements in the work of Grob-Klubmann and Hautsch (2011), Figure 1 (panel b).

[8] See (Engle & Russell, 1998) for an example of U-shaped trading intensity for American stocks.

[9] The settlement costs for a non-institutional investor, a proportional fee charged by the exchange, decreases approximately one basis point for a "day trade" compared to a "Normal trade". Details in http://www.bmfbovespa.com.br/

explain the difference of intraday patterns for trading and news intensity in Figure 3, and also justifies the positive correlation found for news intensity and duration in Table 3. When news intensity is low at the end of the day, the number of trades increases and the duration decreases, which explains the positive correlation found for $pNews_{i,j}$ and $dur_{i,j}$ in Equation (9).

The message from this intraday investigation is clear: a higher number of trades (or smaller duration) does not necessarily imply the existence of news in the market. In fact, closer to the end of the trading day, the increase of trades is not supported by a higher value of news intensity. This means that the trades located at the end of trading hours are not related to the incoming of new information affecting the efficient price of the asset, but due to microstructure frictions in the trading process (i.e., particular motivations of traders such as minimization of overnight risk and trading fee minimization). The proposed method developed in this paper clearly shows such a pattern for the aggregated analysis of the intraday profile of the number of trades (or duration) and news intensity.

# Conclusions

In this paper, we look into the problem of quantifying the presence of news in the financial market based on empirical data. With the support of a theoretical microstructure model, we are able to derive such an estimator and its properties are studied for 20 highly liquid assets of the Brazilian equity market for the time period between 2010 and 2012.

Our first result is that news intensity has a common component across all assets. This is explained by the fact that the assets have similar risk factors such as the country risk. When news regarding the Brazilian economy reaches the market, it affects all of the stocks, thereby, creating a common effect in the vector of news intensity for all assets. Our second result shows that a high likelihood of news is related to high volatility. This is an intuitive result as volatility (or price uncertainty) is related to the incoming of new information into the market. We also see from the regressions that the volume of trades has a negative relationship with news intensity. This can be explained as the bi-product of traders trying to minimize their private information by trading small quantities.

In the paper, we also provide an intraday analysis of news intensity. This part of the study shows that there is a higher intensity of news at the beginning of trading hours, slowing down towards the end of the day. This is a positive result as one can expect higher incoming of overnight news at the beginning of the day. We also show that, at the end of the trading day, number of trades increases, while intensity stays low. We explain this result as the effect of traders avoiding overnight risk and also the particular structure of trading fees in the Brazilian equity market, where trades opened and closed on the same day have lower trading costs.

This paper sets a framework for a novel area of microstructure research, the study of news intensity based on empirical trade data. However, the study used some assumptions that could be modified for more realistic results: in the underlying microstructure model we assumed that the likelihood of news is constant over time. This is counter-intuitive as the intensity of news can change over time and is likely to cluster in the same way as volatility measures. In the empirical part of the paper, we addressed this issue by performing a rolling window in the estimation of news intensity. Future work could look into incorporating such a property directly into the model by addressing the probability of news as a Markov or Garch type of process. There is a clear contribution by inserting the clustering effect directly in the structural model and not in the way that the parameter was estimated from the data.

# References

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, *21*(2), 785–818.

Becker, K. G., Finnerty, J. E., & Friedman, J. (1995). Economic news and equity market linkages between the U.S. and U.K. *Journal of Banking and Finance*, *19*(7), 1191–1210.

Birz, G., & Lott Jr., J. R. (2011). The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking & Finance*, *35*(11), 2791–2800.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Boyd, J. H., Jagannathan, R., & Hu, J. (2001). *The Stock Market's Reaction to Unemployment News: Why Bad News is Usually Good for Stocks* (NBER Working Papers No. 8092). National Bureau of Economic Research, Inc

Carvalho, C., Klagge, N., & Moench, E. (2011). The persistent effects of a false news shock. *Journal of Empirical Finance*, *18*(4), 597 – 615.

Choi, J. Y., Salandro, D., & Shastri, K. (1988). On the Estimation of Bid-Ask Spreads: Theory and Evidence. *The Journal of Financial and Quantitative Analysis*, *23*(2), pp. 219–230.

De Jong, F., & Rindi, B. (2009). *The Microstructure of Financial Markets*. Cambridge University Press, Cambridge, UK.

Dufour, A., & Engle, R. F. (2000). Time and the Price Impact of a Trade. *The Journal of Finance*, *55*(6), 2467–2498. doi:10.1111/0022-1082.00297

Engelberg, J., & Parsons, C. (2011). The Causal Impact of Media in Financial Markets. *The Journal of Finance*, *66*(1), 67–97. doi:10.1111/j.1540-6261.2010.01626.x

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*(4), pp. 987–1007.

Engle, R. F., & Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*, *66*(5), pp. 1127–1162.

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

Garcia, D. (2013). Sentiment during Recessions. *The Journal of Finance*, *68*(3), 1267–1300. doi:10.1111/jofi.12027

Grob-Klubmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, *18*(2), 321 – 340. doi:http://dx.doi.org/10.1016/j.jempfin.2010.11.009

Hasbrouck, J. (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, USA.

Mitchell, M. L., & Mulherin, J. H. (1994). The Impact of Public Information on the Stock Market. *Journal of Finance*, *49*(3), 923–50.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3).

Pearce, D. K., & Roley, V. V. (1985). *Stock Prices and Economic News* (NBER Working Papers No. 1296). National Bureau of Economic Research, Inc.

Rangel, J. G. (2011). Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking and Finance*, *35*(5), 1263–1276.

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance*, *39*(4), 1127–39.

Wasserfallen, W. (1989). Macroeconomics news and the stock market: Evidence from Europe. *Journal of Banking & Finance*, *13*(4-5), 613–626.

Waud, R. N. (1970). Public Interpretation of Federal Reserve Discount Rate Changes: Evidence on the “Announcement Effect.” *Econometrica*, *38*(2), 231–50.

## APPENDIX A - Derivation of $pNews$

We start the derivation of $pNews$ by looking at the price difference equation from the microstructure model given in (1) to (5):

$$\Delta P_t = \epsilon_t + \frac{S}{2}\Delta b_t \qquad \text{(A.1)}$$

From Equation (A.1), we can see that, unconditionally, zero trade price changes can only happen in the following case: $\epsilon_t = -\frac{S}{2}\Delta b_t$. This event can be divided into two separate cases: $\epsilon_t = \Delta b_t = 0$ or $\epsilon_t = -\frac{S}{2}\Delta b_t$ with $\epsilon_t \neq 0$ and $\frac{S}{2}\Delta b_t \neq 0$. Mathematically, assuming now that $S \neq 0$, this translates to the following expression:

$$\begin{aligned} Pr\left(\Delta P_t = 0\right) = Pr(\epsilon_t = 0)Pr\left(\Delta b_t = 0\right) + \\ Pr(\epsilon_t = -\frac{S}{2}\Delta b_t)Pr(\epsilon_t \neq 0)Pr(\Delta b_t \neq 0) \end{aligned} \qquad \text{(A.2)}$$

In Equation (A.2), notice that the left side, $Pr(\Delta P_t = 0)$ is easily available in trade data by simply counting the number of zero trade price differences and dividing it by the number of observations. However, such information does not allow for estimating $pNews$ as the term $Pr(\epsilon_t = -\frac{S}{2}\Delta b_t)Pr(\epsilon_t \neq 0)Pr(\Delta b_t \neq 0)$ is not available since we do not have any information on the distribution function of $\epsilon_t$ or the absolute value of $S$. The solution here is to simplify the problem by restricting the probability space with conditional probabilities. Particularly, we look into the probability of a zero trade price, conditional for the cases where the difference in trade signs is equal to zero ($\Delta b_t = 0$).

Formally, conditioning the probability space of Equation (A.2) to the cases where $\Delta b_t = 0$ implies that $Pr\left(\Delta b_t = 0\right) = 1$ and $Pr(\Delta b_t \neq 0) = 0$. Therefore, the final result is:

$$\begin{aligned} Pr\left(\Delta P_t = 0 \mid \Delta b_t = 0\right) &= Pr(\epsilon_t = 0) \\ &= 1 - pNews \end{aligned} \qquad \text{(A.3)}$$

By isolating parameter $pNews$ in last equation we have:

$$pNews = 1 - Pr\left(\Delta P_t = 0 \mid \Delta b_t = 0\right) \qquad \text{(A.4)}$$