

Estimating the Intensity of News Based on Trade Data

Marcelo Perlin (marcelo.perlin@ufrgs.br)^{*†}

This version: December 19, 2012

WORKING PAPER

Abstract

This paper investigates the problem of identifying the strength of the incoming of news in the financial market. With the support of a microstructure model we are able to derive a simple formula that, based only on trade data, estimates the likelihood of having news for any given tradable asset in a particular time period. The formula can be easily implemented and takes just one input, the probability of a zero trade price difference conditional on the incoming of consecutive same sign trades. In the empirical part of the paper we investigate the properties of this proposed estimator of news intensity for twenty stocks from the Brazilian equity market, covering two full years from 2010 to 2012. The results are very encouraging and consistent across assets. First we find that the strength of news have a common component across all assets. We attribute that to the fact that the incoming of new information regarding the Brazilian economy will affect all stocks. We also see that the likelihood of news is strongly and positively related to volatility of price differences and negatively related to trade volume. The first can be explained by the fact that volatility is a bi-product of news and the second by the presence of traders avoiding the disclosure of private information by trading smaller volumes. In the empirical section we are also able to show that the intensity of news has a intraday pattern, with higher values at the beginning of the trading day and lower values at the end. This result is consistent with the view that the beginning of the trading day is the time when accumulated overnight information reaches the market, therefore increasing news intensity.

^{*}Assistant Professor, Federal University of Rio Grande do Sul (Porto Alegre, Brazil).

[†]The author acknowledges and is very grateful for the financial support from FAPERGS (Fundacao de Amparo a Pesquisa do Estado do Rio Grande do Sul), process number 0854 12-3 (ARD-003/2012).

1 Introduction

The role of news is fundamentally important in the financial literature as it is one of the main factors that move 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 relation of these events to financial prices is that they change the perception of future value by the investors, which in turn update their quotes and change the observed market prices of financial instruments. If most of the investors believe that a recent event will rise (fall) 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. For example, the death of a CEO can be seen as a good news by some and bad news by others. The asymmetry of perception between market participants induces into 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 byproduct of news. Not surprisingly, the empirical effects of news (and volatility) in the financial market soon drawn the attention from the academic community.

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 the result 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 toward stock market's returns.

As we can expect, the last two papers have been followed by a significant amount of research on the case study of impact of news. We emphasize 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, discount rate and real economic. The authors find a strong evidence that unexpected news (surprises) for monetary policy affects the stock market. In more recent years, we can find several studies in the topic. Just to cite a few, we have Mitchell and Mulherin [1994] on the relationship of amount of news announcements and market activity, Barber and Odean [2008] on the effect of media attention over trade prices, Carvalho et al. [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 others research articles¹. These studies provide a body of empirical evidence regarding the effect of news in the financial market under different perspectives.

¹For example see Wasserfallen [1989], Becker et al. [1995], Boyd et al. [2001], Rangel [2011] among others.

In general, the previous literature on news impact can be summarized as a two step research: the incoming of news is identified in the different media (newspaper, internet, among others) 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 a microstructure model and set out to estimate the intensity of the incoming of 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 on that information solely. This is an interesting approach as we treat news as generic information coming to the market and updating investor's belief about the fair price of the asset. 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 type of data, trade prices and trade signs (identities of trade aggressors (buyers or sellers)). As it will be showed 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 twenty 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 that to the incoming of news regarding the Brazilian economy, which will in turn affect all stocks systematically. From our second regression model we report that news intensity has tendency to cluster, with a positive relationship to 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 restricting the sample used in the estimation based on time of the day, we find that in the beginning of the trading hours is when there is a higher value of incoming of news, when comparing to the rest of the trading day. This is an intuitive result as the beginning of the trading hours is exactly the time of the day where is a higher volume of new information to be incorporated into prices. In general, the results from this research shows that the proposed estimator of news intensity has the expected properties when confronted with real data, showing its potential for financial research. Summing up, 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.

2 Estimating the intensity of news

The idea of a microstructure model is to formalize the way in which a price comes to the market. We start the discussion by first criticizing some of the properties of the model in Roll [1984]. We are aware that many models have been developed since², but we focus on this case as it is a benchmark model and we use the critics in order to introduce our own version. The main critics given in this model with respect to the role of news can be easily extended to more recent representations.

The model of Roll [1984] is given by the following set of equations:

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

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

$$\epsilon_t \sim N(0, \sigma^2) \quad (3)$$

$$b_t = \begin{cases} 1 & \text{with 50\% chance} \\ -1 & \text{with 50\% chance} \end{cases} \quad (4)$$

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 ϵ_t . Variable ϵ_t follows a continuous Normal distribution with zero expectation and variance given by σ^2 . The effect of news in the efficient price is the value of ϵ_t . The main intuition in the definition of the price process is that the future efficient price is not predictable since the expectation of price changes are equal to zero, that is $E(\Delta m_t) = 0$. This means that the best predictor for the price in $t+1$ is the price at time t . Parameter S is the spread of this theoretical market and P_t is the trade price at time t , which is a linear function of the trading signs b_t (1 for a buy, -1 for a sell). The trades in this theoretical market are also random, following a 50% chance of a buy or sell.

The attractiveness of Roll's model lies in its simplicity. All the parameters from the model can be easily estimated from trade data (see Frank and Rindi [2009] and Hasbrouck [2007] for details). But, as one can expect, the downside is that it does a poor job in representing the reality of markets. First, markets do not have news all of the time. There are moments in which no new information is arriving. This becomes clear when one compares the properties of the microstructure model against empirical evidence. Notice that given Equation 2, one should never expect a zero price movement ($\Delta P_t = 0$). This is easy to prove analytically by realizing that the price changes of Roll's model will follow:

$$\Delta P_t = \epsilon_t + \frac{S}{2} \Delta b_t \quad (5)$$

²See Hasbrouck [2007] and Frank and Rindi [2009] for a nice review of structural model in microstructure theory.

Since ϵ follows a continuous distribution we can show that $Pr(\Delta P_t = 0) = Pr(\epsilon_t = -\frac{S}{2}\Delta b_t) = 0$. This means that a zero price difference is theoretically impossible if Roll's microstructure model is true. When looking into empirical trade data, this is clearly not the case. From the perspective of an order book structure, when small volume trades consecutively fill a large buy or sell order in one side of the order book, it creates a succession of same price trades and therefore zero trade price differences. Any researcher with access to high frequency trade data can confirm that zero price differences are a common property of the trade price data. For example, for the data used in this study, the percentage of zero price differences across all assets over the course one year can reach as high as 79%, with an average of 63% of all trades. Also, the assumption of Normality for the innovations is highly restrictive and mostly unrealistic toward empirical data. This divergence from the theoretical models and the empirical data cannot be ignored.

Another problem with Roll's microstructure model is with respect to the process of trade signs, which will follow a 50/50 dynamic. The stochastic behavior of the trades implies that the chance of a buy order being followed by another buy order is also 50%³. This is also not found in empirical data, which shows a strong positive autocorrelation in the time series vector of trade signs (identity of aggressor). The result is usually explained by the fact that big traders split their whole order with the objective of minimizing price impact and avoiding the disclosure of private information.

In the paper we suggest an improved version of Roll's microstructure model, which will address the problems disclosed before. The novel model will be given by the following set of equations:

$$m_t = m_{t-1} + \epsilon_t \quad (6)$$

$$P_t = m_t + b_t \frac{S}{2} \quad (7)$$

$$\epsilon_t \sim \begin{cases} = 0 & \text{with probability } 1 - pNews \\ \neq 0 & \text{with probability } pNews \end{cases} \quad (8)$$

$$b_t = \begin{cases} 1 & \text{if } ST_t = 1 \\ -1 & \text{if } ST_t = 2 \end{cases} \quad (9)$$

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

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \quad (10)$$

For Equation 6 the term m_t is the efficient (true) price of an asset, which will follow a random walk representation with innovations given by ϵ_t . Different from the previous case, variable ϵ_t will follow a discrete probability distribution, with zero expectation. Notice that the underlying model has a superficial de-

³Formally $Pr(b_t = 1 | b_{t-1} = 1) = Pr(b_t = 1) = 50\%$.

scription 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 parametrization of the intensity of news in the market, represented by parameter $pNews$, which can take any value between 0 and 1. Sometimes there will be news in the market ($\epsilon_t \neq 0$) and sometimes there won't be ($\epsilon_t = 0$). The higher the value of $pNews$, more intense is the incoming of news, meaning that one can expect a higher number 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 (see Choi et al. [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.⁴ We point out that the simpler model of Roll [1984] is nested with a discrete (and unknown) distribution for ϵ_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). We start the derivation of $pNews$ by looking at the price difference equation from the microstructure model given in 6 to 10:

$$\Delta P_t = \epsilon_t + \frac{S}{2} \Delta b_t \quad (11)$$

From Equation 11 we can see that, unconditionally, zero trade price changes can only happen in the following 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:

$$Pr(\Delta P_t = 0) = Pr(\epsilon_t = 0)Pr(\Delta b_t = 0) + \quad (12)$$

$$Pr(\epsilon_t = -\frac{S}{2} \Delta b_t)Pr(\epsilon_t \neq 0)Pr(\Delta b_t \neq 0) \quad (13)$$

In equation 13 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. But, such information does not allow

⁴From 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 $p_1 = p_2$.

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 ϵ_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 ($\Delta b_t = 0$). With this condition, the second term in the right hand side goes to zero as, given that $\Delta b_t = 0$, then $Pr(\Delta b_t \neq 0) = 0$.

Formally, restricting the probability space of Equation 13, the conditional probability will be given by:

$$Pr(\Delta P_t = 0 | \Delta b_t = 0) = Pr(\epsilon_t = 0) \quad (14)$$

$$= 1 - pNews \quad (15)$$

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

$$pNews = 1 - Pr(\Delta P_t = 0 | \Delta b_t = 0) \quad (16)$$

That is, the probability of news is related to the probability of a zero price movement, conditional on the cases where the trade sign 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 prices differences only when the signs of trades 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.

3 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 3500 financial instruments including stocks and some derivatives from the year of 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.⁵ 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 4075 thousand trades per stock, per day.

The original dataset is stored in different text files and it contains several trading information. This includes session date, instrument ID (e.g, VALE5), trade number, trade price, traded quantity, trade time, trade indicator, order

⁵This is equivalent to UTC minus 3 (2) hours for normal (summer) time.

buy date, sequential order buy number, order sell date and sequential order sell number. Most of the items in the text files are self-explanatory (e.g. trade price, trade volume among others). The items that demand some explanation are the trade indicator and the sequential order buy (sell) number. The trade indicator gives the information of the status of the trade. If by any chance the trade is canceled (e.g. market freeze), this item takes the value of "A" and "X" when trade is a complement of a cancellation (e.g., a canceled buy order that generates two trades). The cases of "A" and "X" are very rare in the data.

The sequential order buy/sell number is an index that address, for a given asset, the order in which the trades happened within a day (the counting is restarted in the beginning of a new day). For example, the first order of the day (either a buy or a sell) will always have a sequential order number of 1. Within the field of sequential order buy/sell number this value is unique for each day and for each stock. Therefore, for all types of orders, there is only one sequential number for each date. It is worth to point out 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. Based on this simple set of rules we create a vector with the trade signs (+1 for a buy, -1 for a sell) for each stock in the dataset.

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 big portion of it. But, 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 canceled order with trade indicator "X" or "A" and the first trade of each day since it has a high duration. Next, Table 1 we show some simple statistics for this data after the pre-processing.

Table 1: Descriptive Statistics of Data

Asset	Average Number of Trades per Day	Average Duration (Sec)	Auto- correlation Diff Price (1st lag)	Auto- correlation Trade Signals (1st lag)
VALE5	8045,69	4,13	-0,29	0,25
OGXP3	6148,38	5,75	-0,31	0,26
BVMF3	5725,78	6,47	-0,34	0,29
ITUB4	5163,68	6,79	-0,27	0,27
PDGR3	4938,18	8,00	-0,31	0,29
GGBR4	4276,05	8,13	-0,27	0,27
BBAS3	4350,45	8,27	-0,25	0,27
BBDC4	4249,78	8,26	-0,25	0,26
ITSA4	4336,82	8,43	-0,34	0,30
USIM5	3899,58	9,04	-0,27	0,27
PETR3	3366,21	10,80	-0,25	0,30
CYRE3	3679,15	9,53	-0,23	0,29
GFSA3	3371,98	11,15	-0,30	0,29
VALE3	3225,00	10,76	-0,22	0,29
MRVE3	3295,84	10,93	-0,25	0,29
CSNA3	2841,29	12,34	-0,23	0,29
RSID3	2691,83	13,60	-0,26	0,31
HYPE3	2582,15	17,39	-0,24	0,29
CIEL3	2689,61	13,60	-0,26	0,28
RDCD3	2640,85	14,25	-0,24	0,28

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) 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 sign 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 fifty million data points. The order of the assets in Table 1 is respecting 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 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 another increase (decrease). 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, Frank 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 corroborate with the arguments presented before for justifying the use of a Markov chain as the stochastic process of trades.

4 Methodology

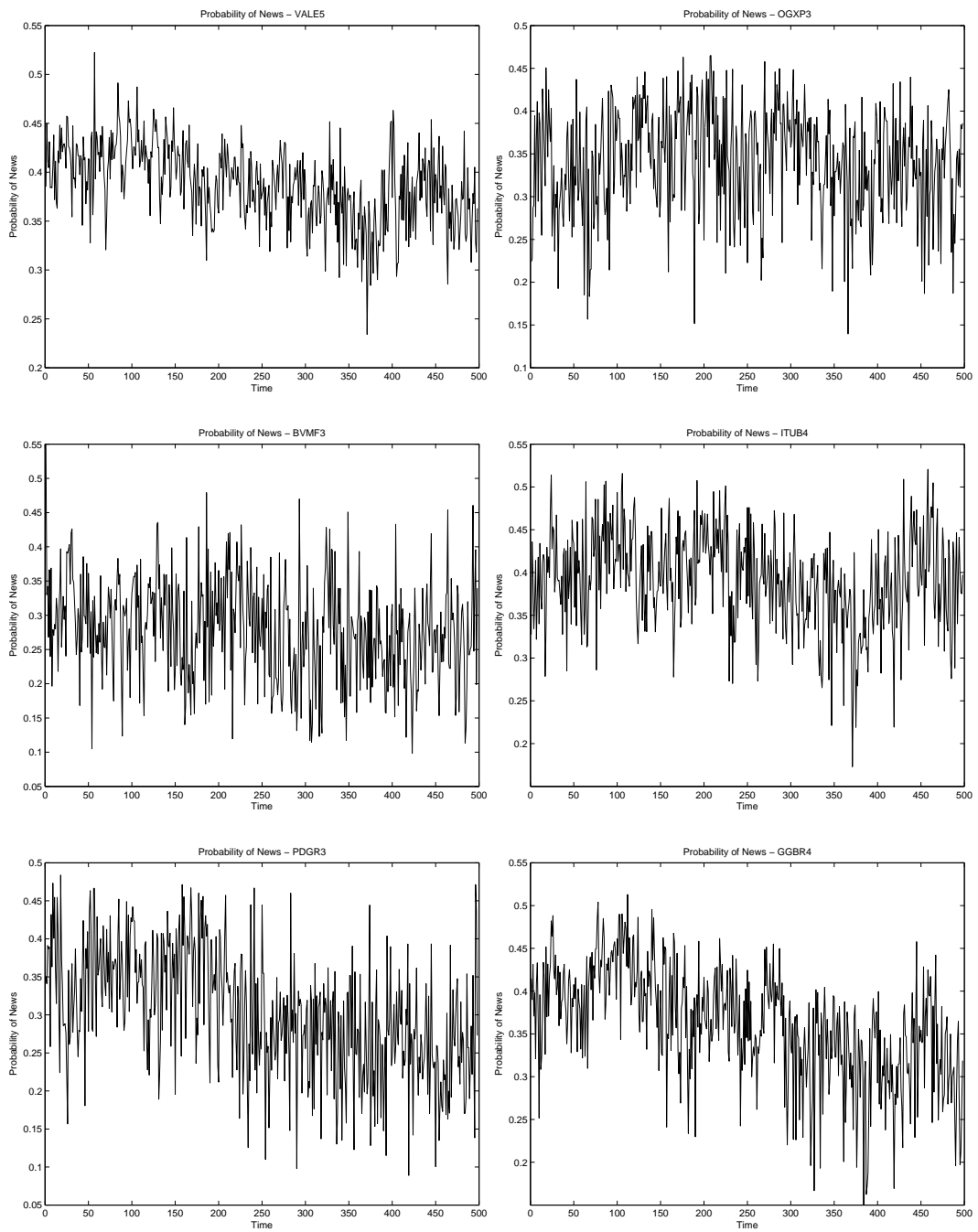
In this section we are interested in studying the properties of the proposed measure of news intensity developed in the first part of the paper. From Equation 16 we can see that it is a measure of the likelihood of news for a particular time period. In order to give it some dynamic, we calculate the 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 directly input it at equation 16. This is given by:

$$pNews_i = 1 - Pr(\Delta P_{i,t} = 0 | b_{i,t} = b_{i,t-1}) \quad (17)$$

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). Next, Figure 1, we show a panel with the time series variation of the likelihood of news for the six most traded assets in the dataset⁶.

⁶We do not include all stocks in order to save space.

Figure 1: Estimated probability of news for the six most traded assets in the research



The first information we notice about the picture in 1 is that the intensity of news varies significantly along the years. For example, for the most traded asset in the dataset, VALE5, it can be as low as 25% and as high as 55%. Also notice that the value of $pNews$ usually ranges around 35% for all stocks. This means that, on average, most of the trades in the dataset are not related to the incoming of new information. This again shows that the assumption that $pNews = 1$ in Roll's model is not realistic when confronting it with real data. We can also see a clustering effect in the dynamics of the estimated intensity of news, where a high (low) value is most likely to be followed by another high (low) value.

The empirical analysis of the paper is divided in 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 of news regarding the Brazilian Market as a whole. This can be either the disclosure of economic reports, a new presidential election, among many other things. 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:

$$pNews_{i,j} = \alpha + \phi pNews_{i,j}^{Agg} + \epsilon_t \quad (18)$$

For equation 18, the term $pNews_{i,j}^{Agg}$ is the average of the probability of news across the assets over time. To explain it better, $pNews$ 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 ϕ 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:

$$pNews_{i,j} = \alpha + \beta_1 pNews_{i-1,j} + \beta_2 Volat_{i,j} + \beta_3 Vol_{i,j} + \beta_4 dur_{i,j} + \epsilon_{i,j} \quad (19)$$

The basic idea in the estimation of 19 is to investigate how the estimated likelihood of news relates to other variables and itself. As we argued in the introduction of the paper, volatility is bi product of news. We also know that volatility have a tendency to cluster⁷. Likewise we should also expect such property for the intensity of news since it is one of the main components of volatility. Following this argument, we add a AR(1) term in order to check for the clustering of news intensity. This property was already found visually in Figure 1, but we add the autoregressive parameter in order to test it formally for our data. We also add contemporaneous volatility, trade volume and duration in our econometric analy-

⁷See Arch/Garch models, Engle [1982] and Bollerslev [1986]

sis. For equation 19 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 is the average duration (time between trades), divided by 100.

5 The Results

We start the presentation for the first model, equation 18.

Table 2: Results from Econometric model, Equation 18

Asset	α	Φ	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
GFS3	-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 $pNews_i = \alpha + \phi pNews_i^C + \epsilon_t$, 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 heteroskedasticity and serial correlation (see Newey and West [1987]). The values with *, ** and *** means statistical significance at 10%, 5% and 1%, respectively.

The results from Table 2 shows 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 ϕ 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 in common a country risk factor and therefore news about Brazil's economy would impact all assets.

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

Table 3: Results from Econometric model, Equation 19

Asset	α	β_1	β_2	β_3	β_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 $pNews_i = \alpha + \beta_1 pNews_{i-1} + \beta_2 Volat_i + \beta_3 Vol_i + \beta_4 dur_i + \epsilon_t$, where $Volat_i$ is the average volatility (standard deviation) of price changes for day i , Vol_i is the average volume of trades and dur_i is the average duration (time between trades). All the standard errors used for statistical testing are robust to heteroskedasticity and serial correlation (see Newey and 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 β_1 is positive and significant for 95% of the cases (nineteen out of twenty). 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 β_2 is positive and significant for all cases of the data. This result provides statistical evidence for the positive relationship between news and volatility of price changes.

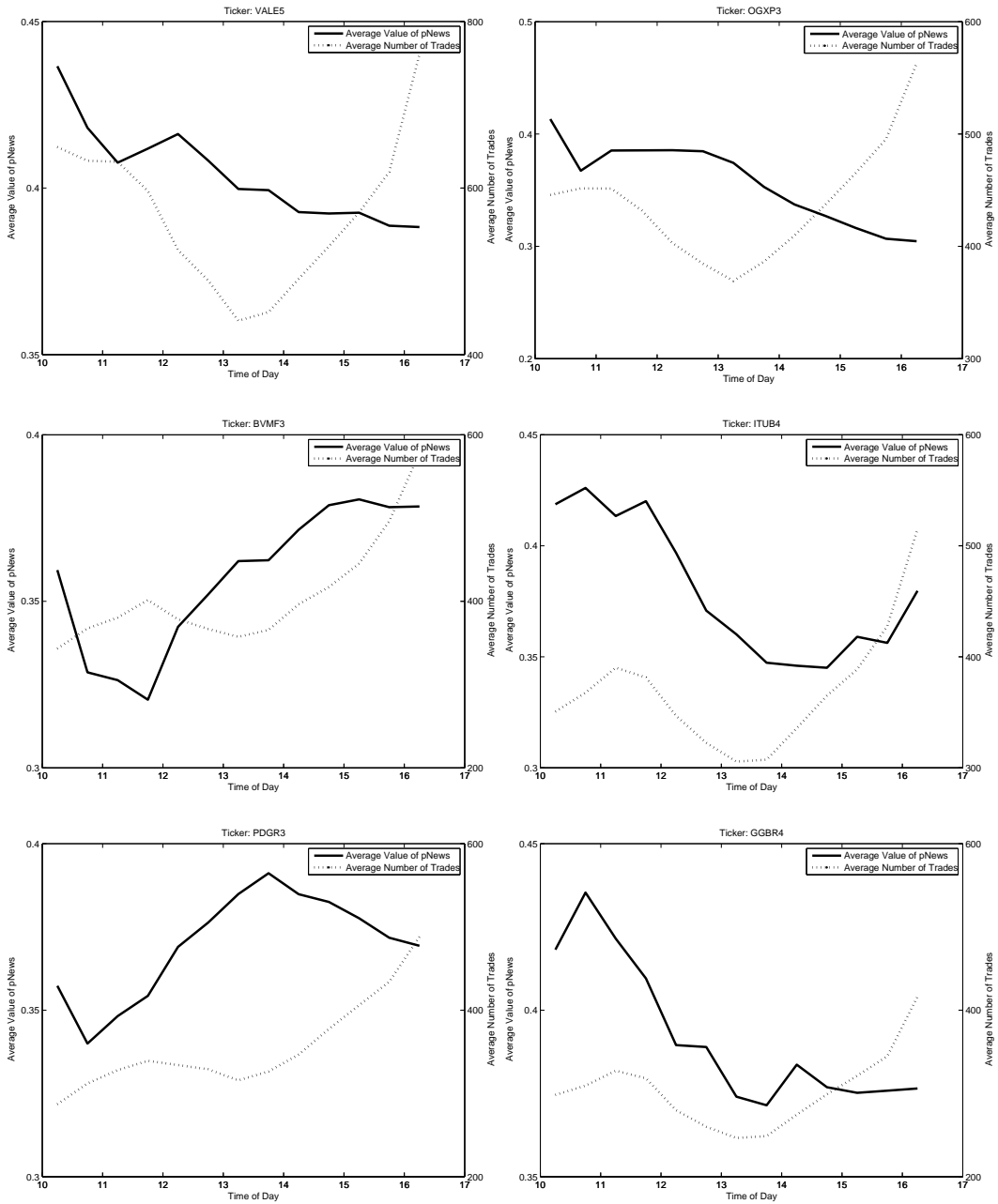
An interesting result is found for the relationship between average volume of trades and the probability of news. We see from Table 3 that the values of β_3 are mostly negative. Out of twenty cases, ten are negative and statistically significant. While this result is not particularly strong, it still does imply 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 β_4 , we find the result that the relationship between duration and news intensity is mostly positive, where eleven cases presented statistically significant coefficients with positive values. This is an interesting, but not intuitive result. It was expected that a higher intensity of news was related to a low duration and higher trading activity. One argument that could explain this result is that the days with higher duration are related to the incoming of scheduled information to the market, such as publication of economic figures (unemployment, etc) and financial statements. In this case the agents (traders) will wait for the release of new information before trading. If this effect is strong enough, the days with higher news intensity will also be the ones with higher duration.

In this study we also look at the intraday profile of the news intensity estimator. This particular investigation is motivated by fact the market have a particular structure of incoming of news due to their opening and closing hours. Next, Figure 2, we show the intraday average of news intensity and number of trades over the period of a trading day for the six most traded assets in the database⁸.

⁸We omit the figure for the 14 other assets in order to save space. But, these can be sent upon request.

Figure 2: Time of the Day, average of news intensity and number of trades for the six most traded assets in the research. These figures were built by first dividing a whole trading day into half hour intervals, calculating the intensity of news with Equation 16 and then averaging it for each time interval. The ticker for each stock is available at the top middle of each individual figure.



The figures in 2 show how the values of news intensity change for each time of the trading day. In order to construct the figure, we divide the hours of the day⁹ between 10:15 and 17:00 into 30 minutes intervals and calculate the value of $pNews$ (see Equation 16) in each one of these intervals. After that, we average the values for each time of the day. The resulting plot is therefore the expected profile (or change) of news intensity during the course of a day.

When looking individually at each panel in 2, a pattern is clear for great majority of the assets, the intensity of news is higher at the beginning of the day. We point out that most of the pictures for the assets not presented in Figure 2 do possess the decaying pattern, with higher average of news intensity at the beginning of the day. This result is intuitive as the opening time of the market is exactly where one can expect a higher volume of new information to arrive. This is due to the closure of the market in the previous day. As the market is close, new information from the overnight period accumulates. When the market re-opens in the next day, this information is then incorporated into prices. It is interesting to see that the proposed measure of news intensity was able to capture such a pattern.

In Figure 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 market microstructure¹⁰. Now, it should be pointed out that a increase of trading intensity is an indication of the incoming of news (see Dufour and Engle [2000]), where traders respond to the new set of available information, adjusting their portfolios according to their expectations and therefore increasing trading intensity.

While the panels 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. While this result is not intuitive, we can present an alternative explanation. We attribute it to the overnight risk and trading costs.

It is well known that traders usually don't 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 of new information, but due to the avoidance of overnight risk, that is, holding a financial position over the night. Also, particular to the case of Brazil, regulation sets that the trading costs are lower if a trade in the stock market is opened and closed in the same day.¹¹ This means that traders in Brazil have not only a risk motivation 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 minimization of costs. This can explain the difference of intraday patterns for trading and news intensity at Figure 2.

⁹These are times (hours) of the day adjusted for summer time, when necessary.

¹⁰See Engle and Russell [1998] for an example of "U" shaped trading intensity for American stocks.

¹¹The settlement costs for non institutional investor, a proportional fee charged by the exchange, decreases approximately one basis point for a "day trade", when comparing to a "Normal trade". Details in <http://www.bmfbovespa.com.br/>.

6 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 twenty 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 therefore 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 in 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 general, the empirical results from the regression models are consistent with main theories in market microstructure.

In the paper we also provide an intraday analysis of news intensity for the six most traded stocks in the dataset. This part of the study shows that there is a higher intensity of news at the beginning of the trading hours, slowing down toward the end of the day. This is an intuitive 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, the trading intensity measured as 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 feed in the Brazilian Equity market, where trades opened and closed at the same day have lower trading cost.

This paper sets a framework for a novel area of microstructure research, the study of news intensity based on empirical trade data. But, the study used some assumptions that could be modified for more realistic results. For example, 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 address this issue by performing a rolling window in the estimation of the news intensity. Future work could look into incorporating such property directly at the model by addressing the probability of news as a Markov or Garch type of process.

Another idea for future research would be to work on identifying the content of news. We could represent the whole process of news as the sum of news in different aspects according to its risk factors. For example, we could separate the whole news as industry related, country related and individual (residual). Possibly a multivariate structural model could be used as the basic framework to identify and measure the intensity of the news for different risk factors. These and other ideas are left for future research.

References

- B. M. Barber and T. Odean. 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, 2008.
- K. G. Becker, J. E. Finnerty, and J. Friedman. Economic news and equity market linkages between the u.s. and u.k. *Journal of Banking and Finance*, 19(7):1191–1210, October 1995. URL <http://ideas.repec.org/a/eee/jbfina/v19y1995i7p1191-1210.html>.
- G. Birz and J. R. Lott Jr. The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking & Finance*, 35(11):2791–2800, November 2011. URL <http://ideas.repec.org/a/eee/jbfina/v35y2011i11p2791-2800.html>.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, April 1986. URL <http://ideas.repec.org/a/eee/econom/v31y1986i3p307-327.html>.
- J. H. Boyd, R. Jagannathan, and J. Hu. The stock market’s reaction to unemployment news: Why bad news is usually good for stocks. NBER Working Papers 8092, National Bureau of Economic Research, Inc, Jan. 2001. URL <http://ideas.repec.org/p/nbr/nberwo/8092.html>.
- C. Carvalho, N. Klagge, and E. Moench. The persistent effects of a false news shock. *Journal of Empirical Finance*, 18(4):597 – 615, 2011.
- J. Y. Choi, D. Salandro, and K. Shastri. On the estimation of bid-ask spreads: Theory and evidence. *The Journal of Financial and Quantitative Analysis*, 23(2):pp. 219–230, 1988. ISSN 00221090. URL <http://www.jstor.org/stable/2330882>.
- A. Dufour and R. F. Engle. Time and the price impact of a trade. *The Journal of Finance*, 55(6):2467–2498, 2000. ISSN 1540-6261. doi: 10.1111/0022-1082.00297. URL <http://dx.doi.org/10.1111/0022-1082.00297>.
- R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):pp. 987–1007, 1982. URL <http://www.jstor.org/stable/1912773>.
- R. F. Engle and J. R. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66(5):pp. 1127–1162, 1998. ISSN 00129682. URL <http://www.jstor.org/stable/2999632>.
- E. Fama, L. Fisher, M. C. Jensen, and R. Roll. The adjustment of stock prices to new information. *International Economic Review*, 10(1):1–21, February 1969. URL <http://ideas.repec.org/a/ier/iecrev/v10y1969i1p1-21.html>.

- D. J. Frank and B. Rindi. *The Microstructure of Financial Markets*. Cambridge University Press, Cambridge, UK, 2009.
- J. Hasbrouck. *Empirical Market Microstructure*. Oxford University Press. New York, 2007.
- M. L. Mitchell and J. H. Mulherin. The impact of public information on the stock market. *Journal of Finance*, 49(3):923–50, July 1994. URL <http://ideas.repec.org/a/bla/jfinan/v49y1994i3p923-50.html>.
- W. K. Newey and K. D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), May 1987. URL <http://ideas.repec.org/a/ecm/emetrp/v55y1987i3p703-08.html>.
- D. K. Pearce and V. V. Roley. Stock prices and economic news. NBER Working Papers 1296, National Bureau of Economic Research, Inc, Apr. 1985. URL <http://ideas.repec.org/p/nbr/nberwo/1296.html>.
- J. G. Rangel. Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking and Finance*, 35(5):1263–1276, May 2011.
- R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4):1127–39, 1984.
- W. Wasserfallen. Macroeconomics news and the stock market: Evidence from europe. *Journal of Banking & Finance*, 13(4-5):613–626, September 1989. URL <http://ideas.repec.org/a/eee/jbfina/v13y1989i4-5p613-626.html>.
- R. N. Waud. Public interpretation of federal reserve discount rate changes: Evidence on the 'announcement effect'. *Econometrica*, 38(2):231–50, March 1970. URL <http://ideas.repec.org/a/ecm/emetrp/v38y1970i2p231-50.html>.