Emerging Markets Review 21 (2014) 21–41

Contents lists available at ScienceDirect

Emerging Markets Review

journal homepage: www.elsevier.com/locate/emr

The adverse selection cost component of the spread of Brazilian stocks

Gustavo Silva Araújo [1], Claudio Henrique da S. Barbedo [*,2], José Valentim M. Vicente [3]

Central Bank of Brazil, Brazil
IBMEC Business School, Brazil

ARTICLE INFO

Article history:
Received 4 November 2013
Received in revised form 7 July 2014
Accepted 22 July 2014
Available online 1 August 2014

Keywords:
Adverse selection cost
Bid–ask spread
Corporate governance

ABSTRACT

This study analyzes the adverse selection cost component embedded in the spreads of Brazilian stocks. We show that it is higher than in the U.S. market and presents an intraday U-shape pattern (i.e., higher at the beginning and at the end of the day). In addition, we investigate the relationships of the adverse selection cost with a firm's characteristics. We find that stocks listed in the highest corporate governance levels do not have the lowest costs. On the other hand, the liquidity of shares, the trade size and the market value of the firm are directly correlated with this cost.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Classic economic theory treats asset prices as a result of Walrasian equilibrium between demand and supply. For financial assets, however, differences in fundamental prices can occur in the short term due to issues related to market microstructure. These issues affect the bid–ask spread of asset prices. In this paper, we study aspects of the adverse selection component embedded in the bid–ask spread of stocks traded in the Brazilian market. In particular, our contribution is to examine the relationship of this component with the size and the time of the trade and to investigate how the connection is between spread and the adverse select component with a firm's characteristics. Finally, we analyze whether stocks

* Corresponding author at: Av. Presidente Wilson, 118 – CEP. 20030-020 - Centro - Rio de Janeiro - RJ, Brazil.
E-mail addresses: 00araujogs@gmail.com (G.S. Araújo), claudio.barbedo@bcb.gov.br (C.H.S. Barbedo), jose.valentim@bcb.gov.br (J.V.M. Vicente).
[1] Tel./fax: +55 2121895503.
[2] Tel.: +55 2121895565; fax: +55 2121895503.
[3] Tel.: +55 2121895672; fax: +55 2121895503.

http://dx.doi.org/10.1016/j.ememar.2014.07.003
1566-0141/© 2014 Elsevier B.V. All rights reserved.

listed in premium corporate governance segments of the Brazilian stock exchange (BM&FBovespa — São Paulo Mercantile, Future and Stock Exchange) have a lower adverse selection cost.

The bid–ask spread is the difference between the highest price that a buyer is willing to pay for an asset (bid price) and the lowest price for which a seller is willing to sell it (ask price). This difference can be seen as a transaction cost to execute an order. In general, the fundamental price of the asset is within this range. There are two classes of models for estimating the bid–ask spread. The first approach, initially proposed by Roll (1984), uses properties of the serial covariance of asset returns. In the second group, the analysis of the spreads relies on regressions in which the independent variable is the trading indicator. This indicator identifies whether the transaction is initiated by a buyer or a seller (Glosten and Harris, 1988). Although the covariance models can be used to determine the spread and its components (see, for example, George et al., 1991 and Stoll, 1989), the methodology based on the direction of the trade is best suited for this purpose.[4]

The bid–ask spread can be attributed to three components: inventory, adverse selection and order processing costs. The inventory cost represents the cost seen by a market maker to provide liquidity to the market. Pioneering work on microstructure, such as Stoll (1978) and Ho and Stoll (1981), consider only the inventory cost in the analysis of the spread. However, other studies point out that the existence of the spread is also due to the adverse selection costs arising with asymmetric information among market participants (see, for example, Glosten and Milgrom, 1985; Lin et al., 2012, and Riordan et al., 2013; Bleaney and Li, 2014 and Li and Xu, 2014). Finally, there are order processing costs such as equipment and personnel (Roll, 1984). Huang and Stoll (1997, HS hereafter) generalize the model based on the trading indicator of Glosten and Harris (1988) by including these three components in the spread. Madhavan et al. (1997) work in the same line, but they do not include the inventory cost.

In this paper, we estimate the components of the bid–ask spread of the major stocks traded on the Brazilian market using the first version of the HS model.[5] In addition, we employ an extension of the HS model to study the patterns of spreads and adverse selection costs as a function of the size and time of the trade. Finally, we implement an extensive research through a series of regressions in order to determine the characteristics of the companies that are correlated with the adverse selection component and the spread. In particular, we analyze the relationship between the adverse selection and corporate governance levels.

In general, corporate governance indexes only take into account aspects of the firms. An innovation of this paper is to study the information asymmetry of the companies (the adverse selection cost), which is a proxy for corporate governance, through the lens of investors' demand rather than using the firm's characteristics. Thus, we can compare the information asymmetry perception of investors with measures of corporate governance built on firm's characteristics.

The sample consists of 52 stocks traded on the BM&FBovespa with data from October 2007 to April 2008. The Brazilian stock exchange had a daily average turnover of $ 3.9 billion in June 2008, which places it as one of the largest stock markets in the world and the largest in Latin America. In order to provide robustness to our results, we split the database in two parts. The first part covers the period from October 18, 2007 to January 18, 2008 and the second is from January 28 to April 24, 2008.

The stock market in Brazil is an order-driven market that has an interesting feature. Although the presence of market makers is allowed and even encouraged, firms with liquid stocks do not have this specialist.[6] Therefore, we limit our study to the stocks without market makers, which represent most of the trading volume.[7] In markets without market makers, the cost of inventory can be neglected. Thus, the two versions of the HS are exactly the models of Glosten and Harris (1988) and Madhavan et al. (1997).

[4] Smith and Whaley (1994) show that the estimates of the spread based on the serial covariance are negatively biased. In addition, Gwilym and Thomas (2002) argue that these estimates may be biased due to noise in the data.

[5] In their article, Huang and Stoll (1997) propose two models known as the first and second models of Huang and Stoll. The difference between the two models is the treatment of the trade autocorrelations.

[6] Market makers on the Brazilian Stock Exchange have the commitment of being in the market daily with firm buy and sell offers for a given number of assets. By registering the offers, they set publicly quoted prices.

[7] The shares in our sample represent 88.5% of Ibovespa, the main index of Brazilian stock Exchange. The stocks that integrate Ibovespa's theoretical portfolio represent more than 80% of the financial value and the number of trades on BM&FBovespa.

In May 2008, foreign investors are the largest players in the Brazilian stock market (about 35%), followed by regular people (28%), institutional investors (26%), financial institutions (7%) and others (4%). Trading shares on the Brazilian stock exchange require the intermediation of a brokerage firm. The stocks are traded on Mega Bolsa, an electronic trading system that allows buying or selling orders to be registered via computer terminals. Since 1999 local clients can also use a Home Broker system (provided by brokerage firm) which allows the investors to register orders directly on Mega Bolsa.

An important methodological contribution of our work concerns the procedure used to obtain the sequence of trade initiations. Rather than estimate it by approximated methods, as in other studies, we determine this sequence from the nature of the trade actually carried out. This approach is only possible because our database contains, besides tick-to-tick trading prices, the time of buy and sell offers. The approximated methods of estimating the sequence of initiations are accurate to about 80% (see, for example, Omrane and Welch, 2013). Therefore, our results are not affected by the bias of the initiation sequence estimation.

Our results can be summarized as follows. The cost of adverse selection in the BM&FBovespa is, on average, larger than in the U.S. market. In terms of patterns, we note that the spread and the adverse selection component increase with the size of the trade and are higher at the beginning of the day. The adverse selection is also high at the end of the day, i.e., it has a U-shaped related to trading time. Moreover, the most liquid stocks present the lowest spreads and adverse selection costs. Regarding the firm size, we find that bigger companies have the lowest adverse selection cost. Volatility, unlike in the U.S. market, is not significant to explain the spread. The fact that a stock is listed in premium corporate governance listing segments does not affect the spread and the adverse selection component. Therefore governance corporate levels built using only characteristics of the firm may not capture the investor's perception. Since the results show that adverse selection cost (a proxy of disclosure) is lower as the liquidity is higher, a possible way to circumvent that problem is to give more importance to liquidity in the assessment of corporate governance.

This article is organized as follows. Section 2 describes the database used. In Section 3, we present the models. In Section 4, we analyze the spread and the adverse selection estimates and their patterns regarding trade size and hours. Section 5 discusses the variables that are related to the adverse selection cost and its relationship with corporate governance. In this section we also present a robustness check of our findings. In Section 6, we offer our concluding remarks.

2. Sample and database treatment

2.1. Sample

The database was built by the BM&FBovespa specifically for this study. It is composed of three distinct parts. Parts one and two contain information about the buy and sell orders, which includes the time stamp of each trade up to the millionth of a second, the stock code, the identifier of the order, the validity of the order, and the volumes and the price. Part three contains information about effectively traded stocks. This part consists of date, stock code, price, trading volume, date–time and identifier of the buy order, and date–time and identifier of the sell order.

We use two periods to check the robustness of our results. The first period is from October 18, 2007 to January 18, 2008 and the second is from January 28 to April 24, 2008.[8] Both periods contain 60 days. This sample size is consistent with several other studies.[9] We chose these time periods because they comprise both uptrends and downtrends for the stocks in our sample. Thus, we believe that the periods chosen did not influence the results.[10] Furthermore, since April 2008 there were no changes in the microstructure of the market that might jeopardize the conclusions of our study.

[8] There is a small gap between the first and the second period in order to remove the split of the Companhia Siderúrgica Nacional share of the data.

[9] See, for instance, Ahn et al. (2002) and De Winne and Majois (2003).

[10] The second period is a pre-crisis period when the international crisis had already started to show its effects on some stocks. It is recommended not to use microstructure models based on measures of central tendency in the whirlwind of severe crises, because of the increase in herding behavior.

The BM&FBovespa was traditionally an order-driven market. However, since 2002, it has allowed the presence of market makers to provide liquidity for some stocks. Thus, in the Brazilian stock market, some shares have market makers. However, there is only one book of limit orders for each stock. Since in Brazil, the most liquid stocks in the BM&FBovespa do not have market makers, we work only with shares without these specialists.[11]

2.2. Database treatment

We refine the sample by selecting trades that simultaneously meet the following criteria: a) the trade does not have any type of change afterwards; b) the trade is not canceled; and c) the trade occurs in the range of 10:05 am to 16: 55 h (from 11:05 to 17:55 during the daylight savings time). The first and second criteria ensure the validity of the operation. The third aims to remove the prices formed on opening and closing auctions.

By analyzing the intraday trading, we identify the origin of each transaction as arising from a buyer or a seller (buyer-initiated or seller-initiated) by matching information from the buy order database, sell order database and trading order database. If the buy order occurs after the sell order, it is considered as initiated by the buyer and $Q = +1$. Otherwise, the operation is considered as initiated by the seller and $Q = -1$. There are also trades in which the buy order occurs at the same second as the sell order and, in this case, $Q = 0$. With this information we build a database composed of date, hour, volume, price changing related to the previous trade and the indicator of initiation (Q).[12]

Unlike other studies that use methods based on the direction of the trade, this study precisely identifies transactions as either buy or sell. Works in the U.S. market use techniques based on checking if the price is below or above the last trade price. The techniques commonly used for this purpose are the quote method, the tick test and the LR method (Lee and Ready, 1991). However, these methods are not accurate. For example, Ellis et al. (2000) find, using data from NASDAQ, accuracies of 76%, 78% and 80% for the quote method, the tick test and the LR method, respectively. Lu and Wei (2009), using data from the Taiwan Stock Exchange, report accuracies of 93% for the quote method, 74% for the tick test and 97% for the LR method. Omrane and Welch (2013) find that the approximated methods of estimating the sequence of initiations are accurate to about 80%. In the Brazilian market, Perlin et al. (2014) report accuracy of 72% for the tick test.

Furthermore we perform a second treatment, because some records in our database refer to the same order. Consider, for instance, that there are two limit orders for selling, one of 200 shares at $40.00 and another of 100 shares at $40.30, and both orders have the lowest sell prices in the limit order book. An order to buy 300 shares at $40.30 generates two trades in the BM&FBovespa database. We modify the database so that this order only generates one trade of 300 shares and price of$ 40.10 (average price per share of the trade).

Appendix A contains the selected stocks of the study. The sample includes 4,128,997 trades from October 18, 2007 to January 18, 2008 and 4,517,530 from January 28 to April 24, 2008. There are 4,127,019 (47.73%) classified as buyer initiated, 3,989,656 (46.14%) classified as seller initiate and 529,852 trades (6.13%) not identified either way.

3. Spread and adverse selection models

By nature, inventory costs exist only in quote-driven markets, where specialists have the institutional obligation to supply liquidity continuously (De Jong and Rindi, 2009). On the other hand, adverse selection and order-processing costs may exist in any financial market.

Since the Brazilian stock market is order driven and we only study companies without market makers, there is no need to use models that include the inventory cost. Then we use the first model of Huang and Stoll (1997) with inventory cost equal to zero. This model is based on the nature of the trade indicator. We also used a generalized version of the first model of Huang and Stoll (1997) to detect possible patterns in

[11] Unlike the Brazilian exchange, stocks in the NYSE have a market maker.
[12] The first trade of each day is removed because we should not use the price changing relative to the previous day trade.

the spread and in the adverse selection cost. These models are all well known but for completeness and to develop notation we provide a brief summary here.[13]

3.1. First model of Huang and Stoll (1997)

Trade indicator models assume that bid and ask prices are the result of competition among all players in the market (Glosten, 1987). There is no assumption that the bid and ask quotes represent the same individual, i.e., these types of model can be used for stocks without market makers.

Let p^* be the value of the stock if all agents have access to inside information. Suppose that the risk of inside information is not priced. In this case, the 'true' price of the stock, based on all common-knowledge information (H), is $p = \mathrm{E}[p^*|H]$.

Assuming that investors generally have only common-knowledge information, we can define the functions $a(.)$ and $b(.)$:

$$a(x) = \mathrm{E}[p^*|H, \text{“investor buys at x”}].$$

$$b(y) = \mathrm{E}[p^*|H, \text{“investorsells aty”}].$$

The functions $a(x)$ and $b(y)$ describe how the common-knowledge are updated to include the information about the previous trade.

Let $Z_A = a(A) - p$ and $Z_B = p - b(B)$, where A and B are respectively the ask and bid prices. Then $Z_A + Z_B$ is the part of the spread due to the belief that there are informed investors. We can write A and B as:

$$A = a(A) + C_A = p + Z_A + C_A$$

$$B = b(B) – C_B = p – Z_B – C_B,$$

where C_A and C_B are order processing costs. Then the spread $S = A - B$ is given by $Z_A + Z_B + C_A + C_B$.

Set Q_{n+1} as an indicator variable that is $+1$ if the trade $n+1$ is buyer initiated and -1 if this trade is seller initiated. Define also ε_{n+1} as the revision of the true price (p_n) due to the arrival of new public information between the trades n and $n+1$. Thus, the true price is

$$p_{n+1} = p_n + \varepsilon_{n+1} + Z_{n+1} Q_{n+1}, \tag{1}$$

where $Z_{n+1} = Z_A$ if $Q_{n+1} = +1$ and $Z_{n+1} = Z_B$ if $Q_{n+1} = -1$. Note that there are two innovations in the true price, one due to public information and the other due to the previous transaction. The trade price is

$$\hat{p}_{n+1} = p_{n+1} + CQ_{n+1}, \tag{2}$$

where $C = C_A$ if $Q_{n+1} = +1$ and $C = C_B$ if $Q_{n+1} = -1$, $C_A, C_B > 0$.

With Eqs. (1) and (2) and the assumptions that $Z_A = Z_B = Z$, $C_A = C_B = \mathrm{C}$ and that Z and C are constants, we arrive at Eq. (3) from which we estimate the spread s and the adverse selection cost (as a percentage of the spread) α. The steps to obtain Eq. (3) are in Appendix B.2.

$$\Delta\hat{p}_n = \varepsilon_{\mathrm{n}} + \alpha\frac{S}{2}Q_{n-1} + \frac{S}{2}\Delta Q_n. \tag{3}$$

[13] For details, see Glosten (1987), Glosten and Harris (1988) and Huang and Stoll (1997).

The models in this section are estimated by GMM (generalized method of moments), which imposes weak assumptions about the distributions. This is an important issue since ε_n can include rounding errors because the trade prices are discrete. The estimation results of this study are robust to several conditions of orthogonality, for the presence of conditional heteroskedasticity and serial autocorrelation.

3.2. Generalized model of Huang and Stoll (1997)

We also use a generalization of the first model of Huang and Stoll (1997), which allows the determination of spread and the information asymmetry patterns. Again, the model present here neglects the inventory cost because we are studying stocks without market makers. In this section, we describe, as an example, how the patterns related to transaction size, in terms shares traded, are obtained. The same procedure can be used to analyze patterns of any other variable, such as the trade period, which is also studied in this article. By fixing $0 < k < j$, the volume of stocks transacted in trade $n + 1$ are classified as s, m or l, according to the following rule:

s-volume $\leq k$ shares
m-volume between k and j shares
l-volume $\geq j$ shares.

The estimated equation is

$$\Delta \hat{p}_n = \varepsilon_n + \frac{S^s}{2}\Delta D^s_n + \alpha^s \frac{S^s}{2} D^s_{n-1} + \frac{S^m}{2}\Delta D^m_n + \alpha^m \frac{S^m}{2} D^m_{n-1} + \frac{S^l}{2}\Delta D^l_n + \alpha^l \frac{S^l}{2} D^l_{n-1} \tag{4}$$

where:

$D^s_{n+1} = Q_{n+1}$, if the volume $\leq k$ shares and 0 otherwise.
$D^m_{n+1} = Q_{n+1}$, if the volume is between k and j shares and 0 otherwise.
$D^l_{n+1} = Q_{n+1}$, if the volume $\geq j$ shares and 0 otherwise.
s^s — spread when the volume $\leq k$ shares.
s^m — spread when the volume is between k and j shares.
s^l — spread when the volume $\geq j$ shares.

Appendix B.3 shows the steps to obtain Eq. (4). To estimate the intraday patterns of the spread and of the asymmetric information component, one needs only to change the definitions of s, m and l by the periods of the day.

4. Results of the spread and the adverse selection patterns

4.1. Spread and adverse selection component

Appendix A shows the spread and adverse selection costs from January 28 to April 24, 2008 estimated by the first model of Huang and Stoll (1997).[14] The values of these two variables are consistent with the literature: all stocks present positive adverse selection costs and, from 52 stocks, 50 present costs lower than 100% of the spread. In addition, 46 stocks present a spread greater than the minimum tick (one cent of real). The stocks with spreads lower than the minimum tick have the lowest prices. The fact that some stocks have spread lower than the minimum tick may be explained by the presence of trades in the same second (joint initiated). Since there are about 6% of trades classified as jointly initiated, the spread estimates are likely slightly underestimated in relation to the effective spread.

The most liquid stocks, such as Vale (VALE5) and Petrobras (PETR4),[15] present the lowest spreads in terms of percentage of the average price. The percentage spread of PETR4 is the smallest, but the absolute

[14] The results from October 18, 2007 to January 18, 2008 are not shown because they are similar.
[15] The stocks of these companies represent about 30% of the BM&FBovespa Index.

Table 1
Descriptive statistics of spreads and adverse selection costs (AS) estimated by the first model of Huang and Stoll (1997) from October 18, 2007 to January 18, 2008. Adverse selection costs are presented as a percentage of the spread and spread is presented in cents of real and in percentage of the price.

	Average spread	Average spread	AS
	(Cents of real)	(% of the average price)	(% of the spread)
1st quartile	2.05	0.07%	57.21%
Average	3.82	0.11%	64.60%
Median	3.00	0.09%	64.08%
3rd quartile	5.13	0.13%	71.91%

Table 2
Descriptive statistics of spreads and adverse selection costs (AS) estimated by the first model of Huang and Stoll (1997) from January 28 to April 24, 2008. Adverse selection costs are presented as a percentage of the spread and spread is presented in cents of real and as a percentage of the price.

	Average spread	Average spread	AS
	(Cents of real)	(% of the average price)	(% of the spread)
1st quartile	1.72	0.06%	57.58%
Average	3.31	0.10%	66.61%
Median	2.61	0.09%	65.03%
3rd quartile	4.13	0.12%	75.34%

spread of VALE5 is the smallest. This result occurs because if a stock has a lower price (in this case, VALE5), it tends to have a higher spread as a percentage of the price because there is a minimum tick. VALE5 and PETR4 are also the stocks that have the lowest adverse selection cost. Although PETR4 is the most traded stock over the period, it presents higher adverse selection cost than VALE5.

Tables 1 and 2 show the descriptive statistics of spread and adverse selection estimated by the first model of Huang and Stoll (1997) from October 18, 2007 to January 18, 2008 and from January 28 to April 24, 2008, respectively. The results are similar for both periods, but the adverse selection component is slightly larger for the second period and the average spread (as a percentage of average price) is slightly higher in the first interval (in Section 5 we investigate the relationship between spread and adverse selection). Compared to the international literature, particularly for the U.S., the adverse selection in the Brazilian market is larger (see, for instance, Huang and Stoll, 1997; Glosten and Harris, 1988; and Tannous et al., 2013). This result may be due to the fact the U.S. market is much more liquid and analyzed.

4.2. Spread and adverse selection patterns

In this section, we implement the generalized model of Huang and Stoll (1997), presented in Section 3.2, to verify spread and adverse selection patterns according to the trade size and intraday hours.

Table 3
Descriptive statistics of spreads and adverse selection costs (AS) estimated by the first model of Huang and Stoll (1997) from October 18, 2007 to January 18, 2008 according to the trade size. Adverse selection costs are presented as a percentage of the spread and spread is presented in cents of real and as a percentage of the price.

	Small trades			Medium trades			Large trades		
	Average spread	Average spread	AS	Average spread	Average spread	AS	Average spread	Average spread	AS
	(Cents of real)	(% of the average price)	(% of the spread)	(Cents of real)	(% of the average price)	(% of the spread)	(Cents of real)	(% of the average price)	(% of the spread)
1st quartile	1.85	0.06%	38.65%	1.96	0.06%	54.39%	2.50	0.08%	78.24%
Average	3.93	0.11%	53.70%	3.58	0.11%	61.58%	4.42	0.13%	83.95%
Median	3.24	0.09%	48.84%	2.72	0.08%	61.20%	3.71	0.11%	84.19%
3rd quartile	5.73	0.15%	66.52%	4.85	0.12%	68.61%	5.35	0.15%	90.47%

Table 4
Descriptive statistics of spreads and adverse selection costs (AS) estimated by the first model of Huang and Stoll (1997) from January 28 to April 24, 2008 according to the trade size. Adverse selection costs are presented as a percentage of the spread and spread is presented in cents of real and as a percentage of the price.

	Small trades			Medium trades			Large trades		
	Average spread	Average spread	AS	Average spread	Average spread	AS	Average spread	Average spread	AS
	(Cents of real)	(% of the average price)	(% of the spread)	(Cents of real)	(% of the average price)	(% of the spread)	(Cents of real)	(% of the average price)	(% of the spread)
1st quartile	1.72	0.06%	45.93%	1.60	0.06%	51.46%	2.27	0.08%	76.46%
Average	3.47	0.11%	63.54%	3.13	0.10%	61.45%	3.77	0.12%	83.68%
Median	2.67	0.09%	52.19%	2.37	0.08%	60.06%	3.24	0.11%	82.02%
3rd quartile	4.81	0.13%	68.72%	3.79	0.10%	72.20%	4.53	0.14%	90.16%

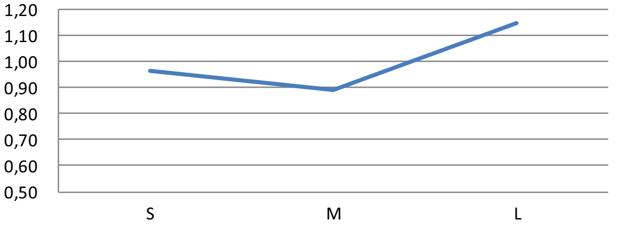

Fig. 1. Average spread normalized by trade size from October 18, 2007 to January 18, 2008. These normalized variables are the spread divided by the average of the three stock size variables.

4.2.1. Trade size patterns

We define trade size as the number of shares traded. As the stock prices of our database are quite different, we do not choose specific values for classifying a trade as small, medium or large. Instead, for every stock, we classify as small trades those below the 20th percentile of the range during the entire period, as medium trades those between the 20th and 80th percentiles, and as large trades those above the 80th percentile.

Tables 3 and 4 present the descriptive statistics of the spreads and the adverse selection costs (as a percentage of the spread) estimated by the HS model from October 18, 2007 to January 18, 2008 and from January 28 to April 24, 2008, respectively. It can be seen that the percentage spread and the adverse selection component are higher for the large trades. Moreover, the adverse selection component is higher as the size of the trade increases.

To analyze the spread and adverse selection patterns according to the trade size, we normalize these variables. The normalized variables are composed by the spread or the adverse selection of a specific size divided by the average of all size variables of the stock. For instance, the normalized spread for small trades of PETR4 is the spread of small trades of PETR4, s^s, divided by the average of the spreads, $(s^s + s^m + s^l) / 3$, of PETR4. The same applies to the adverse selection component. Figs. 1–4 show the average of each size among the stocks.

Figs. 1 and 3 show that the normalized spreads are similar for the two periods.[16] The spreads are lower for the medium trades, i.e., they have a U-shaped pattern. As in the U.S. market (see, e.g., Huang and Stoll, 1997), the spread is higher for trades generated by larger buy and sell orders.

Figs. 2 and 4 examine the adverse selection component.[17] These figures and Tables 3 and 4 show that this component is higher as the trade size increases and is very much higher for larger orders. This result is

[16] For robustness, we calculate the correlations of the three normalized values of the spread. Because there are 52 stocks, we have 1326 different correlations. For the first period, 877 (or 66.14%) of these correlations are above 0.5, and for the second period, 856 (or 64.56%) of these correlations are above this value.

[17] For robustness, we calculate the correlations of the three normalized values of the adverse selection component. Because there are 52 stocks, we have 1326 different correlations. For the first period, 925 (or 69.76%) of these correlations are above 0.5, and for the second period, 823 (or 62.07%) of these correlations are above this value.

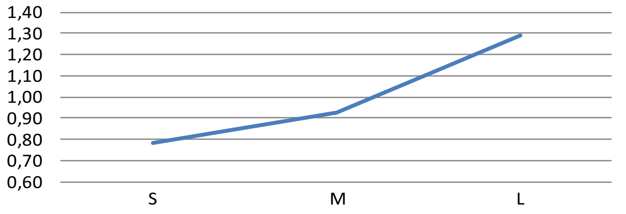

Fig. 2. Average adverse selection costs normalized by trade size from October 18, 2007 to January 18, 2008. These normalized variables are the adverse selection costs divided by the average of the three stock size variables.

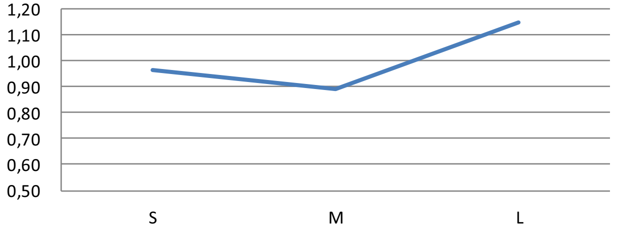

Fig. 3. Average spread normalized by trade size from January 28 to April 24, 2008. These normalized variables are the spread divided by the average of the three stock size variables.

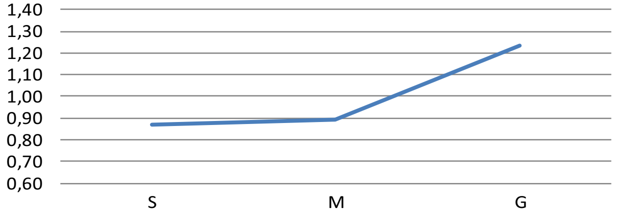

Fig. 4. Average adverse selection costs normalized by trade size from January 28 to April 24, 2008. These normalized variables are the adverse selection costs divided by the average of the three stock size variables.

similar to Ahn et al. (2002) who studied the Tokyo Stock Exchange, which as the Brazilian market is a limit order market, and Zhao et al. (2013) who studied the Taiwan Stock Exchange. Our results suggest that big trades have a higher probability of being initiated by an insider. On the other hand, Barclay and Warner (1993) and Huang and Stoll (1997) find that medium trades contain more asymmetric information than large trades on the NYSE.

4.2.2. Intraday trading pattern

The regular trading on the BM&FBovespa occurs for 7 h. We investigate how intraday spread and adverse selection values change hourly. Figs. 5–8 are similar to Figs. 1–4, but, instead of trade sizes, they present the trade hours. Hour three, for instance, refers to the third hour of trading.[18] Figs. 5 and 6 are from the first period and 7 and 8 from the second. Figs. 5 and 7 show the spread pattern and 6 and 8 show the adverse selection component.

[18] We take into account the daylight saving time.

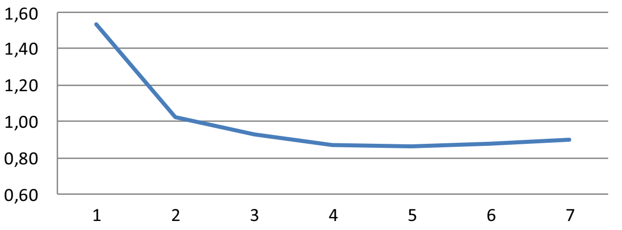

Fig. 5. Average spread normalized by trading hours from October 18, 2007 to January 18, 2008. These normalized variables are the spread divided by the average of the seven trading hour variables.

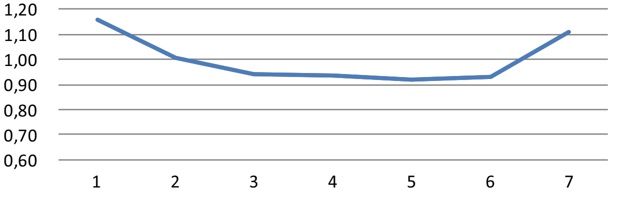

Fig. 6. Average adverse selection costs normalized by trading hours from October 18, 2007 to January 18, 2008. These normalized variables are the adverse selection costs divided by the average of the seven trading hour variables.

The patterns for the two periods are similar. The spread decreases quickly in the first 2 h, becoming almost flat after the third hour. This result is different from Chung (1999), Lehman and Modest (1994) and Madhavan et al. (1997), all of whom report spreads increasing at the end of the day.

The adverse selection pattern is U-shaped, i.e., at the beginning and at the end of trading there is higher asymmetry. This result is commonly found in the literature (see, for example, Ahn et al., 2002) but contradicts others, like Madhavan et al. (1997), who observe that asymmetry does not increase at the end of trading. Furthermore, the literature is unanimous in reporting higher asymmetry in early trading, possibly due to the fact that in the hours before trading an investor can obtain information, but cannot make any trades.

5. Variables related to the adverse selection component

In this section, we evaluate how the adverse selection component estimated by the model of Huang and Stoll (1997) is associated with stocks' trading characteristics. We also analyze the variables related to the spread for comparison purposes. Initially, we evaluate the spreads following the seminal article by Demsetz (1968). We run a regression with the average spread of the period as the dependent variable:

$$Spread_i = c_1 + c_2 Risk_i + c_3\,\mathrm{Price}_i + c_4 Qty_i + \varepsilon_i, \tag{5}$$

where $Risk_i$ is a risk measure of asset i, $Price_i$ is the average trading price of asset i and Qty_i is the average daily quantity of asset i traded in the period.

Following Demsetz, we first adopt the standard deviation of daily returns as the risk measure. Unlike other papers' results, where higher volatility implies a higher spread, volatility is not significant at 5% in both periods. Table 5 shows the least squares estimation of the spread for the first and second periods. We also consider another measure of risk, the daily returns beta of each stock (a proxy for market risk), estimated on the past 60 months. This variable also is not significant in both periods.[19] It seems that stock

[19] We do not show a table with the variable beta because the results are similar to those in Table 5.

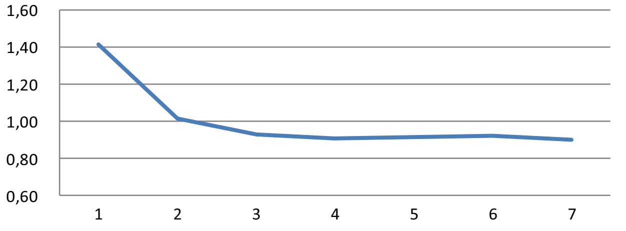

Fig. 7. Average spread normalized by trading hours from January 28 to April 24, 2008. These normalized variables are the spread divided by the average of the seven trading hour variables.

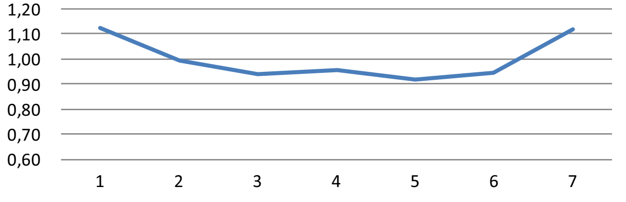

Fig. 8. Average adverse selection costs normalized by trading hours from January 28 to April 24, 2008. These normalized variables are the adverse selection costs divided by the average of the seven trading hour variables.

risk measures do not affect the spread of Brazilian stocks without market makers. Volatility is particularly important for the spread of stocks with market makers, as highlighted by Prucyk (2005), and this may be the reason why the volatility is significant in the literature.

In both periods, the spread presents a strong positive relationship with the stock price and a negative relationship with the quantity of asset traded, consistent with literature.[20] The results show that an increase in the price of R$ 1 (one real) represents an increase of about R$ 0.06 in the spread. An increase of 1000 stocks in trading volume contributes to a fall of R$ 0.75 in the spread.[21]

We did the same regression 12, but now using the spread in percentage of the average price instead of the spread in absolute values. The results are presented in Table 6. All variables except the risk (standard deviation) are statistically significant. The increase of 1000 shares traded decreases in 0.0189% the spread (0.0148% in the second period). With a decrease of $ 10 on average share price, the spread rate increases 0.007%. The sign of the coefficient of the average price is negative because of the existence of the minimum tick.[22]

To determine which variables are related to the adverse selection cost, we regress this variable, as a percentage of the spread, using the same independent variables of Eq. (5). Eq. (6) shows the regression and Table 7 presents the results for both periods:

$$AS_i = c_1 + c_2\mathrm{Risk_i} + c_3\mathrm{Price_i} + c_4\mathrm{Qty_i} + \varepsilon_i. \tag{6}$$

Differently of the spread, the adverse selection component does not have a significant relationship with average stock price. The higher the trading volume, lower the adverse selection cost. Note that the adjusted R^2 decreases substantially in relation to the regression of the spread.

[20] See Benston and Hagerman (1974) and Barbedo and Lemgruber (2008).
[21] The Brazilian Real/US$ exchange rate was around 1.75 in the period of the sample.
[22] Suppose that the stock price is R$ 20 and the spread is R$ 0.01. This spread can only increase because R$ 0.01 is the minimum tick. If the price falls to R$ 10 and the spread does not change, the spread as a percentage of average price increases.

Table 5
Regression of the spread with the average daily quantity of the stock, price and risk as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	3.91187	0.0326	
Quantity	−0.00075	0.0014	
Risk	−3.92385	0.2948	
Average price	0.06223	0.0000	
Adjusted R2	0.437	F-test	0.0000
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	2.48451	0.0630	
Quantity	−0.00065	0.0002	
Risk	−2.29227	0.4134	
Average price	0.07006	0.0000	
Adjusted R2	0.505	F-test	0.0000

Table 6
Regression of the spread (in percentage of the price) with the average daily quantity of the stock, price and risk as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.0020160	0.0002	
Risk	−0.0009040	0.3947	
Average price	−0.0000070	0.0349	
Quantity	−1.8900000E−07	0.0041	
Adjusted R2	0.291	F-test	0.0008
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.0020040	0.0000	
Risk	−0.0012710	0.1168	
Average price	−0.0000069	0.0273	
Quantity	−1.4800000E−07	0.0024	
Adjusted R2	0.279	F-test	0.0000

We did the same regression, but instead of using the average daily quantity, as in Demsetz (1968), we use the liquidity ratio (LR), which also takes into account the trading volume.[23] Eq. (7) presents the regression and Table 8 shows the results for both periods.

$$AS_i = c_1 + c_2\mathrm{Risk_i} + c_3\mathrm{Price_i} + c_4\mathrm{LR_i} + \varepsilon_i \tag{7}$$

Average price and risk still have no significant relationship with the adverse selection component. The adjusted R^2 increases compared to the previous regression. The adverse selection component decreases with the liquidity ratio.

[23] The formula of the liquidity ratio is 100 ∗ p / P ∗ squared-root (n / N ∗ v / V), where p is the number of days that there was at least a trade with the stock within the chosen period; P is the total number of days in the selected period, n is the number of trades of the stock within the chosen period, N is the number of trades of all stocks within the chosen period, v is the traded volume (in monetary units) of the stock within the chosen period, and V is the traded volume of all stocks within the chosen period.

Table 7
Regression of the adverse selection cost with the average daily quantity of the stock, price and risk as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.5192560	0.0000	
Risk	0.3226810	0.1082	
Average price	0.0006080	0.3201	
Quantity	−0.0000347	0.0047	
Adjusted R2	0.116	F-test	0.030
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.6636820	0.0000	
Risk	0.1050590	0.6298	
Average price	0.0001310	0.8746	
Quantity	−0.0000366	0.0055	
Adjusted R2	0.105	F-test	0.040

We also use, instead of the liquidity ratio, the variable size (market value of the firm) and the ratio between the number of firm shares traded in the period and number of shares outstanding (traded/out). Table 9 shows the results.

Only size is significant from the two new variables. Among the proxies for liquidity chosen, the significant variables are the average quantity, the firm size and the liquidity ratio. To avoid multicollinearity problem, we run the adverse selection component against each one of these variables separately and by the criteria AIC (Akaike, 1974) and SIC (Schwarz, 1978); the model with the variable liquidity ratio is the one that better fits.

Next, we add the variable spread, as a percentage of the average price to Eq. (7). We do not use the spread estimated by the HS model, since we would have an independent and dependent variables coming from the same regression. The variable used is the average closing spread of the stock. Eq. (8) presents the regression and Table 10 shows the results. The adverse selection component has, as expected, a positive relation with the spread.

$$AS_i = c_1 + c_2\mathrm{Risk_i} + c_3\mathrm{Price_i} + c_4\mathrm{LR_i} + c_5\mathrm{Spread_i} + \varepsilon_i \tag{8}$$

A point not analyzed in the international literature concerns the influence of corporate governance issues in the spread and the adverse selection component. Improvement of corporate governance practices is an

Table 8
Regression of the adverse selection cost with the liquidity ratio (LR) of the stock, price and risk as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.5133860	0.0000	
Risk	0.3055280	0.1249	
Average price	0.0006390	0.2979	
LR	−0.0231730	0.0045	
Adjusted R2	0.118	F-test	0.029
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.6589070	0.0000	
Risk	0.0839900	0.6983	
Average price	0.0002490	0.7659	
LR	−0.0291930	0.0048	
Adjusted R2	0.109	F-test	0.036

Table 9
Regression of the risk, average of the stock price, size, ratio between the number of shares of the firm traded in the period and number of shares outstanding of the firm (traded/out) and risk as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.47654	0.0000	
Risk	0.40629	0.0723	
Average price	0.00114	0.1124	
Size	−8.39E−10	0.0037	
traded/out	−0.13490	0.1964	
Adjusted R2	0.117	F-test	0.0422
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.63145	0.0000	
Risk	0.16721	0.4815	
Average price	0.00085	0.3739	
Size	−9.00E−10	0.0098	
traded/out	−0.16847	0.1161	
Adjusted R2	0.09111	F-test	0.0748

important strategy recommended by several authors, such as La Porta et al. (1998) and Dai et al. (2013), as well as multilateral organizations, such as OECD, to reduce information uncertainty. Leal and Carvalhal-da-Silva (2007) demonstrate that markets price the quality of a firm's corporate governance practices. This may be the reason why firms would be interested to incur this costly signaling about their behavior.

The BM&FBovespa has adopted an interesting approach to deal with this potential costly signaling. In 2001 it introduced differentiated corporate governance levels: three premium trading segments with specific disclosure and corporate governance practice requirements beyond what is mandatory by corporate law in Brazil. The three premium listing segments are Level 1 (L1), which requires more disclosure than the traditional segment, Level 2 (L2), which requires everything in L1 plus an assortment of corporate governance practices, and finally the New Market (NM), which is equal to L2 with the additional requirement excluding companies from using nonvoting shares.

Table 10
Regression of the adverse selection cost with the liquidity ratio (LR) of the stock, closing spread, average price and risk as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.62123	0.0000	
Risk	0.07456	0.5639	
Average price	0.00030	0.4582	
LR	−0.02139	0.0320	
Spread	10.72680	0.0236	
Adjusted R2	0.279	F-test	0.0048
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.57914	0.0000	
Risk	0.08441	0.6842	
Average price	0.00051	0.5313	
LR	−0.02084	0.0450	
Spread	11.51562	0.0254	
Adjusted R2	0.247	F-test	0.0087

Table 11
Regression of the adverse selection cost with the liquidity ratio (LR) of the stock, closing spread, price, risk and the premium listing dummies as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.36332	0.0030	
Risk	0.34966	0.0899	
Average price	0.00127	0.0592	
LR	−0.01820	0.0299	
Spread	8.57672	0.0574	
L1	0.07543	0.0993	
L2	0.05071	0.5066	
NM	0.06399	0.1969	
Adjusted R2	0.145	F-test	0.0489
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.57848	0.0000	
Risk	0.02741	0.8993	
Average price	0.00064	0.4621	
LR	−0.01933	0.0694	
Spread	11.13555	0.0386	
L1	0.00450	0.9319	
L2	0.03808	0.6787	
NM	0.05783	0.3273	
Adjusted R2	0.156	F-test	0.0398

Premium listings based on corporate governance practices may foster investor confidence when trading. Investors may feel that there is a lower likelihood that they will be at the other end of an insider initiated transaction in which they may end up losing money due to information asymmetry.

In order to observe the relationship between information uncertainty and premium listing segments, we include dummies for these corporate governance levels in Eq. (8). Eq. (9) shows the regression.

$$AS_i = c_1 + c_2\mathrm{Risk}_i + c_3\mathrm{Price}_i + c_4\mathrm{LR}_i + c_5\mathrm{Spread}_i + c_6L1_i + c_6L2_i + c_6NM_i + \varepsilon_i \tag{9}$$

where $L1$, $L2$ and NM are dummies for the corporate governance levels of each stock i. If some dummy is significant, it means that the level of corporate governance affects the adverse selection component. Moreover, if the sign of the dummy coefficient is negative, the level of corporate governance contributes to the reduction of asymmetry. Table 11 presents the results for the first and second periods. The adjusted R^2

Table 12
Regression of the adverse selection cost with the liquidity ratio (LR) of the stock and closing spread as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.67594	0.0000	
LR	−0.02128	0.0349	
Spread	8.43965	0.0285	
Adjusted R2	0.195	F-test	0.0015
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.63360	0.0000	
LR	−0.01927	0.0521	
Spread	11.11536	0.0266	
Adjusted R2	0.208	F-test	0.0012

Table 13
Regression of the adverse selection cost with the liquidity ratio (LR) of the stock, closing spread and the dummy ADR as dependent variables.

Period 1: October 18, 2007 to January 18, 2008			
Variables	Coefficient	P-value	
Intercept	0.65574	0.0000	
LR	−0.02354	0.0378	
Spread	8.39859	0.0301	
ADR	0.01616	0.7251	
Adjusted R2	0.186	F-test	0.0048
Period 2: January 28 to April 24, 2008			
Variables	Coefficient	P-value	
Intercept	0.62839	0.0000	
LR	−0.02051	0.0503	
Spread	11.10785	0.0280	
ADR	0.01616	0.6825	
Adjusted R2	0.194	F-test	0.0038

Table 14
Fixed effect regression of the adverse selection cost with the liquidity ratio (LR) of the stock, closing spread, average price and risk as dependent variables. The subscript 2 refers to the second period of the sample (January 28 to April 24, 2008) and 1 to the first (October 18, 2007 to January 18, 2008).

Variables	Coefficient	P-value	
Intercept	0.02214	0.1916	
$\text{Risk}_2 - \text{risk}_1$	0.30693	0.0962	
$(\text{Average price})_2 - (\text{average price})_1$	0.00011	0.8998	
$\text{LR}_2 - \text{LR}_1$	−0.07835	0.0173	
$\text{Spread}_2 - \text{spread}_1$	−0.02103	0.1886	
Adjusted R2	0.030	F-test	0.2505

decreases and neither of these dummies is significant. This suggests that the adverse selection cost is not affected by the transparency level and corporate governance practices.

As the number of firms at level 2 is small, we grouped all stocks listed in the corporate governance levels in a dummy called GC. When we ran the regression, the dummy variable remains insignificant. We also ran the regression 15 with only the dummies related to the segments of governance, and the variables still remained non-significant. Finally, we ran the adverse selection component against GC and this dummy is also not significant. Thus, corporate governance appears to be unrelated with the adverse selection component.

Table 12 shows the results of the regression only with the statistically significant variables:

Adverse selection component decreases with liquidity. For each unit increase in the liquidity ratio, we have a reduction of approximately 2% of the adverse selection component. Adverse selection component increases with spread. For an increase in 1% of the spread, the adverse selection increases by 9.8%.

From the results above and based on the hypothesis that Brazilian stocks listed on the NYSE present a lesser adverse selection component, because firms are obliged to follow international accounting standards and disclosure requirements, we run Eq. (10):

$$AS_i = c_1 + c_1 \text{LR}_i + c_2 \text{Spread}_i + c_3 \text{ADR}_i + \varepsilon_i \tag{10}$$

where ADR is a dummy that characterizes whether the firm has ADR type II or III.[24]

[24] The ADRs of types II and III are those which are traded on the New York Stock Exchange (NYSE). Other types of ADR are traded over the counter.

From the 52 shares of our sample, 22 (42.33%) are classified as ADR type II or III and seven of these shares are not included in the premium listing segments of the Brazilian Stock Exchange. Table 13 presents the results of regression 10. The dummy variable coefficient is not significant, i.e., there is no relationship between ADR stocks and information uncertainty.

Regarding the adverse selection component, the empirical results of this section suggest that although the premium listing segments require enhanced corporate governance practices, these practices do not change the cost related to the perception of insider trading. This component is correlated to the liquidity according to Table 11. Liquidity means a significant amount of buyers and sellers arriving at a transaction price and at a given time. It reduces the information uncertainty and provides quality price discovery.

Price discovery reflects the ability of the market to find the fundamental value (O'Hara, 2001), which refers to the underlying features of the firm. To prevent manipulating and self-dealing, stock exchanges should monitor disclosure requirements, for instance, throughout corporate governance standards. However, they should also monitor how quickly prices adjust to fundamental values. This study contributes to highlight the importance of setting a higher weight to stock liquidity in the governance levels classification procedure.

5.1. Robustness with a fixed effects regression

In the previous section we present evidences that the adverse selection cost of Brazilian stocks is related to liquidity. In order to check the robustness of this result, we run a fixed effect regression to control for omitted variables concerning idiosyncratic features of the stocks. Note that our data can be analyzed as a panel. The time dimension is given by the two subperiods of the sample while the cross-sectional unit of observation is represented by each stock. The fixed effect equation, with the same variables of regression 8, is given by:

$$\begin{aligned} AS_{i,2}-AS_{i,1} = {} & c_1 + c_2\left(\text{Risk}_{i,2}-\text{Risk}_{i,1}\right) + c_3\left(\text{Price}_{i,2}-\text{Price}_{i,1}\right) + c_4\left(\text{LR}_{i,2}-\text{LR}_{i,1}\right) \\ & + c_5\left(\text{Spread}_{i,2}-\text{Spread}_{i,1}\right) + u_i \end{aligned} \tag{11}$$

where the subscripts 1 and 2 denote the two subperiods of the sample.

Table 14 presents the outcome of the fixed effect regression. Note that the coefficient c_4 remains negative and significant confirming that liquidity is directly correlated to the adverse selection cost. In addition, the other explanatory variables are again not statistically significant.

6. Conclusion

In this article we analyzed the adverse selection component embedded in the bid–ask spread of stocks traded in the Brazilian market. First, we studied the patterns with respect to the trade size and the time of the transaction. Second, we investigated the relationship of the spread and the adverse selection component with a large set of firm's features. The results show that adverse selection cost in Brazil is higher than in the U.S. market. It is also higher at the beginning and at the end of the day and positively related to trade size. The adverse selection cost is not lower for stocks traded in a segment requiring higher levels of corporate governance. This component is instead affected by liquidity. Thus, the stock liquidity must have a special role in the definition corporate governance levels.

A natural extension of this work would be to estimate the components of the spread of Brazilian stocks that have a market maker and compare the results with the conclusions of this work. One should be careful with these comparisons because the stocks with a market maker in Brazil have little liquidity. Another suggestion for future work is to determine the sources of spread and adverse selection cost variations during the day found in this work with the methodologies of Madhavan et al. (1997) or Amihud (2002).

Appendix A

Spreads and adverse selection costs (AS) of the stocks estimated by the first model of Huang and Stoll (1997) from October 18, 2007 to January 18, 2008. Adverse selection costs are presented as a percentage of the spread and spread is presented in cents of real and in percentage of the price.

Stock	Average spread (Cents of real)	Average spread (% of the average price)	AS (% of the spread)	AS (Cents of real)
AMBV4	6.37	0.05%	62.27%	3.97
ARCZ6	0.68	0.06%	103.20%	0.70
BBAS3	1.70	0.06%	46.12%	0.78
BBDC4	1.89	0.04%	59.66%	1.13
BNCA3	5.04	0.21%	74.04%	3.73
BRAP4	3.31	0.08%	67.47%	2.23
BRTO4	1.50	0.08%	84.38%	1.26
BRTP3	6.09	0.12%	75.29%	4.59
BTOW3	6.02	0.09%	67.35%	4.05
CESP6	3.27	0.09%	59.51%	1.95
CGAS5	8.47	0.20%	62.74%	5.31
CLSC6	5.90	0.14%	75.47%	4.45
CMIG4	1.80	0.06%	64.18%	1.16
CNFB4	1.20	0.23%	57.20%	0.69
CPLE6	2.31	0.08%	90.42%	2.09
CRUZ3	4.68	0.10%	83.55%	3.91
CSAN3	4.06	0.15%	69.43%	2.82
CSNA3	3.06	0.05%	59.26%	1.81
CYRE3	2.53	0.10%	69.22%	1.75
DURA4	3.57	0.10%	65.17%	2.32
ELET6	2.42	0.09%	71.03%	1.72
EMBR3	1.11	0.06%	80.28%	0.89
FFTL4	15.78	0.18%	55.54%	8.76
GFSA3	2.18	0.07%	75.46%	1.65
GGBR4	2.18	0.04%	59.05%	1.29
GOAU4	6.83	0.09%	55.20%	3.77
GOLL4	2.37	0.08%	70.16%	1.66
ITAU4	1.69	0.04%	57.71%	0.97
ITSA4	0.89	0.09%	41.54%	0.37
LAME4	1.36	0.10%	59.27%	0.80
LREN3	3.86	0.12%	72.45%	2.80
NATU3	2.22	0.12%	75.79%	1.68
NETC4	1.71	0.09%	61.29%	1.05
PCAR4	1.95	0.06%	71.96%	1.40
PETR4	2.55	0.03%	46.75%	1.19
PRGA3	2.98	0.07%	73.08%	2.18
RAPT4	2.68	0.18%	69.73%	1.87
RDCD3	3.55	0.13%	50.15%	1.78
SBSP3	2.19	0.06%	109.85%	2.41
SDIA4	0.91	0.09%	51.02%	0.47
SLCE3	6.33	0.26%	84.43%	5.35
SUZB5	3.71	0.14%	48.25%	1.79
TCSL4	0.61	0.10%	54.91%	0.34
TLPP4	4.34	0.09%	89.91%	3.90
TNLP4	2.80	0.07%	63.69%	1.78
UGPA4	5.02	0.08%	53.66%	2.69
UNIP6	0.54	0.33%	34.87%	0.19
USIM5	5.40	0.05%	64.22%	3.47
VALE5	1.73	0.04%	44.13%	0.76
VCPA4	3.04	0.06%	77.88%	2.37
VIVO4	0.90	0.09%	79.55%	0.71
WEGE3	2.73	0.13%	64.89%	1.77

Appendix B. Details of spread and adverse selection models

B.1. Description of variables

Here we describe variables that are used in Appendix B.2 and Appendix B.3.

A is the ask price.
B is the bid price.
p^* is the value of the stock if all agents have access to inside information.
p is the 'true' price of the stock, based on all common-knowledge information (H), is $p = \mathrm{E}[p^*|H]$.
$\hat{p}_{n+1}$ is the trade price of the trade $n + 1$.
$a(x) = \mathrm{E}[p^*|H$, "investor buys at x"].
$b(y) = \mathrm{E}[p^*|H$, "investor sells at y"].
Z refers to adverse selection costs.
$Z_A = a(A) - p$.
$Z_B = p - b(B)$.
C refers to order processing costs.
$C_A = A - a(A)$.
$C_B = b(B) - B$.
S refers to the spread.
$S = \mathrm{A} - \mathrm{B}$.
Q_{n+1} is an indicator variable that is $+1$ if the trade $n + 1$ is buyer initiated and -1 if this trade is seller initiated.
ε_{n+1} is the revision of the true price (p_n) due to the arrival of new public information between the trades n and $n + 1$.
α is proportions of $\frac{S}{2}$ due to Z. $\alpha = Z/\frac{S}{2}$.
π is the proportions of $\frac{S}{2}$ due to C. $\pi = Z/\frac{S}{2}$.

B.2. First model of Huang and Stoll

Here we show the steps to get Eq. (3) from Eq. (2).

If we assume $Z_A = Z_B$ and $C_A = C_B$, we have $S = A - B = 2(Z + C)$ or $\frac{S}{2} = Z + C$. Moreover, Z is positive because when an investor buys at price A, E [$p^*|H$, "The investor buys at A"] is higher than $\mathrm{E}[p^*|H]$, i.e., $a(A) > p$ and $Z_A > 0$.

Suppose Z and C are constants. Let α and π be proportions of $\frac{S}{2}$ due to Z and C, respectively. As $Z_A = Z_B$, α is also the proportion of the spread (S) due to asymmetric information ($2Z$).

Since $\alpha = Z/\frac{S}{2}$, the true price is

$$p_{n+1} = p_n + \varepsilon_{n+1} + \alpha\frac{S}{2}Q_{n+1} \tag{12}$$

and the trade price for the trade $n + 1$ is

$$\hat{p}_{n+1} = p_{n+1} + \pi\frac{S}{2}Q_{n+1}. \tag{13}$$

Taking the first difference of this equation, we obtain:

$$\Delta\hat{p}_{n+1} = \Delta p_{n+1} + \pi\frac{S}{2}\Delta Q_{n+1}.$$

Substituting Δp_{n+1} from the first equation in this equation, we have

$$\Delta\hat{p}_{n+1} = \varepsilon_{n+1} + \alpha\frac{S}{2}Q_{n+1} + \pi\frac{S}{2}\Delta Q_{n+1}$$

$$\Delta\hat{p}_{n+1} = \varepsilon_{n+1} + \alpha\frac{S}{2}Q_{n+1} + (1-\alpha)\frac{S}{2}\Delta Q_{n+1}$$

$$\Delta\hat{p}_n = \varepsilon_n + \alpha\frac{S}{2}Q_{n-1} + \frac{S}{2}\Delta Q_n.$$

We estimate α and s from this equation. Alternatively, in terms of Q_{n-1} and Q_n, we have

$$\Delta\hat{p}_n = \varepsilon_n - (1-\alpha)\frac{S}{2}Q_{n-1} + \frac{S}{2}Q_n.$$

B.3. Generalized model of Huang and Stoll (1997)

Here we show the steps to get Eq. (4).
Eq. (12) from Appendix B.2 also can be rewritten as:

$$p_{n+1} = p_n + \varepsilon_{n+1} + \alpha^s\frac{S^s}{2}D^s_{n+1} + \alpha^m\frac{S^m}{2}D^m_{n+1} + \alpha^l\frac{S^l}{2}D^l_{n+1}\,, \tag{14}$$

where:

$D^s_{n+1} = Q_{n+1}$, if the volume $\leq k$ shares and 0 otherwise.
$D^m_{n+1} = Q_{n+1}$, if the volume is between k and j shares and 0 otherwise.
$D^l_{n+1} = Q_{n+1}$, if the volume $\geq j$ shares and 0 otherwise.
s^s — Spread when the volume $\leq k$ shares.
s^m — Spread when the volume is between k and j shares.
s^l — Spread when the volume $\geq j$ shares.

We also can write Eq. (13) from Appendix B.1 as follows

$$\hat{p}_{n+1} = p_{n+1} + (1-\alpha^s)\frac{S^s}{2}D^s_{n+1} + (1-\alpha^m)\frac{S^m}{2}D^m_{n+1} + \left(1-\alpha^l\right)\frac{S^l}{2}D^l_{n+1}.$$

Taking the first difference of the last equation, we obtain:

$$\Delta\hat{p}_{n+1} = \Delta p_{n+1} + (1-\alpha^s)\frac{S^s}{2}\Delta D^s_{n+1} + (1-\alpha^m)\frac{S^m}{2}\Delta D^m_{n+1} + \left(1-\alpha^l\right)\frac{S^l}{2}\Delta D^l_{n+1}.$$

Using Δp_{n+1} from Eq. (14), we have:

$$\Delta\hat{p}_{n+1} = \varepsilon_{n+1} + \frac{S^s}{2}\Delta D^s_{n+1} + \alpha^s\frac{S^s}{2}D^s_n + \frac{S^m}{2}\Delta D^m_{n+1} + \alpha^m\frac{S^m}{2}D^m_n + \frac{S^l}{2}\Delta D^l_{n+1} + \alpha^l\frac{S^l}{2}D^l_n,$$

or

$$\Delta\hat{p}_n = \varepsilon_n + \frac{S^s}{2}\Delta D^s_n + \alpha^s\frac{S^s}{2}D^s_{n-1} + \frac{S^m}{2}\Delta D^m_n + \alpha^m\frac{S^m}{2}D^m_{n-1} + \frac{S^l}{2}\Delta D^l_n + \alpha^l\frac{S^l}{2}D^l_{n-1}$$

Alternatively, we can write the last equation as:

$$\Delta\hat{p}_n = \varepsilon_n + (\alpha^s-1)\frac{S^s}{2}D^s_{n-1} + \frac{S^s}{2}D^s_n + (\alpha^m-1)\frac{S^m}{2}D^m_{n-1} + \frac{S^m}{2}D^m_n + \left(\alpha^l-1\right)\frac{S^l}{2}D^l_{n-1} + \frac{S^l}{2}D^l_n.$$

References

Ahn, H., Cai, J., Hamao, Y., Ho, R., 2002. The components of the bid–ask spread in a limit-order market: evidence from the Tokyo Stock Exchange. J. Empir. Financ. 9, 399–430.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series. J. Financ. Mark. 5 (1), 31–56.
Barbedo, C., Lemgruber, E., 2008. The effect of bid–ask prices on Brazilian options implied volatility: a case study of telemar call options. J. Emerg. Mark. 13, 18–27.
Barclay, Michael J., Warner, Jerold B., 1993. Stealth trading and volatility. J. Financ. Econ. 34, 281–305.
Benston, G., Hagerman, R., 1974. Determinants of bid–asked spreads in the over-the-counter market. J. Financ. Econ. 1 (4), 353–364.
Bleaney, M., Li, Z., 2014. A new spread estimator. University of Nottingham School of Economics Discussion Papers 14/01 (available in https://www.nottingham.ac.uk/economics/documents/discussion-papers/14-01.pdf).
Chung, D., 1999. Bid–ask spread components in an order-driven environment. J. Financ. Res. 22 (2), 227–246.
Dai, L., Fu, R., Kang, J.-K., Lee, I., 2013. Internal corporate governance and insider trading. Working PaperErasmus Univ (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103523).
De Jong, F., Rindi, B., 2009. The Microstructure of Financial Markets. Cambridge University Press.
De Winne, R., Majois, C., 2003. A comparison of alternative spread decomposition models on Euronext Brussels. Brussels Econ. Rev. 46 (Winter).
Demsetz, H., 1968. The cost of transacting. Q. J. Econ. 82, 33–53.
Ellis, K., Michaely, R., O'Hara, M., 2000. The accuracy of trade classification rules: evidence from Nasdaq. J. Financ. Quant. Anal. 35, 529–551.
George, T.J., Kaul, G., Nimalendran, M., 1991. Estimation of the bid–ask spreads and its components: a new approach. Rev. Financ. Stud. 4, 623–656.
Glosten, L.R., 1987. Components of the bid–ask spread and the statistical properties of transaction prices. J. Financ. 42, 1293–1307.
Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid–ask spread. J. Financ. Econ. 21, 123–142.
Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14, 71–100.
Gwilym, O., Thomas, S., 2002. An empirical comparison of quoted and implied bid–ask spreads on futures contracts. J. Int. Financ. Mark. Inst. Money 12, 81–99.
Ho, T., Stoll, H.R., 1981. Optimal dealer pricing under transactions and return uncertainty. J. Financ. Econ. 9, 47–73.
Huang, R.D., Stoll, H.R., 1997. The components of the bid–ask spread: a general approach. Rev. Financ. Stud. 10, 995–1034.
La Porta, R., Lopez-De-Silanes, F., Shleifer, A., Vishny, R., 1998. Law and finance. J. Polit. Econ. 106, 1113–1155.
Leal, R.P.C., Carvalhal-da-Silva, A.L., 2007. Corporate governance and value in Brazil (and Chile). In: Chong, A., Lopez-de-Silanes, F. (Eds.), Investor Protection and Corporate Governance — Firm Level Evidence Across Latin America. Stanford University Press, Palo Alto, California, pp. 213–287.
Lee, M.C., Ready, M., 1991. Inferring trade direction from intraday data. J. Financ. 46, 733–746.
Lehman, B.N., Modest, D.M., 1994. Trading and liquidity on the Tokyo Stock Exchange: a bird's eye view. J. Financ. 44, 951–984.
Li, B., Xu, Y., 2014. Behavioral heterogeneity and financial markets: locked/crossed markets under informationally efficient pricing. Working PaperWashington University (available in https://economics.wustl.edu/tags/other).
Lin, Y., You, S., Huang, M., 2012. Information asymmetry and liquidity risk. Int. Rev. Bus. Res. Pap. 8 (1), 112–131.
Lu, Y., Wei, Y., 2009. Classification of trade direction for an equity market with price limit and order match: evidence from the Taiwan Stock Market. Invest. Manage. Financ. Innov. 6, 135–147.
Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? A transaction-level analysis of NYSE stocks. Rev. Financ. Stud. 10, 1035–1064.
O'Hara, M., 2001. Designing markets for developing countries. Int. Rev. Financ. 2, 205–215.
Omrane, W., Welch, R., 2013. Tick test accuracy in foreign exchange ECN markets. Working PaperBrock University (available in http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2014-Rome/papers/EFMA2014_0302_fullpaper.pdf).
Perlin, M., Brooks, C., Dufour, A., 2014. On the performance of the tick test. Q. Rev. Econ. Financ. 54 (1), 42–50.
Prucyk, Brian R., 2005. Specialist risk attitudes and the bid–ask spread. Financ. Rev. 40 (2), 223–255.
Riordan, R., Storkenmaier, A., Wagener, M., Zhang, S., 2013. Public information arrival: price discovery and liquidity in electronic limit order markets. J. Bank. Financ. 37 (4), 1148–1159.
Roll, R., 1984. A simple implicit measure of the effective bid–ask spread in an efficient market. J. Financ. 39, 1127–1139.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.
Stafford, J.T., 1987. The Share-Owner's Guide: How to Invest Profitable and Safely in Shares. Woodhead-Faulkner, Cambridge.
Smith, T., Whaley, R.E., 1994. Estimating the effective BID/ASK spread from time and sales data. J. Futures Markets 14 (4), 437–455.
Stoll, H.R., 1978. The supply of dealer services in securities markets. J. Financ. 33, 1133–1151.
Stoll, H.R., 1989. Inferring the components of the bid–ask spread: theory and empirical tests. J. Financ. 44, 115–134.
Tannous, G., Wang, J., Wilson, C., 2013. The intraday pattern of information asymmetry, spread, and depth: evidence from the NYSE. Int. Rev. Financ. 13 (2), 215–240.
Zhao, Y., Cheng, L., Chang, C., Ni, C., 2013. Short sales, margin purchases and bid–ask spreads. Pac. Basin Financ. J. 24, 199–220.