Adverse selection costs, trading activity and price discovery in the NYSE: An empirical analysis

Roberto Pascual [a,1], Alvaro Escribano [b,*], Mikel Tapia [c,2]

[a] *Departamento de Economía y Empresa, Universidad de las Islas Baleares, 07071 Palma de Mallorca, Baleares, Spain*
[b] *Departamento de Economía, Universidad Carlos III de Madrid, C/Madrid, 126, 28903 Getafe, Madrid, Spain*
[c] *Departamento de Economía de la Empresa, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain*

Abstract

This paper studies the role that trading activity plays in the price discovery process of a NYSE-listed stock. We measure the expected information content of each trade by estimating its permanent price impact. It depends on observable trade features and market conditions. We also estimate the time required for quotes to incorporate all the information content of a particular trade. Our results show that price discovery is faster after risky trades and also at the extreme intervals of the session. The quote adjustment to trade-related shocks is progressive and this causes risk persistency and unusual short-term market conditions.
© 2002 Elsevier B.V. All rights reserved.

[*] Corresponding author. Tel.: +34-91-624-9854; fax: +34-91-624-9329.
E-mail addresses: rpascual@uib.es (R. Pascual), alvaroe@eco.uc3m.es (A. Escribano), mtapia@emp.uc3m.es (M. Tapia).
[1] Tel.: +34-971-17-2845; fax: +34-971-17-3426.
[2] Tel.: +34-91-624-9639; fax: +34-91-624-9608.

1. Introduction

A central issue of the financial microstructure literature is the analysis of the information content of trades. From a theoretical viewpoint, the stochastic process of prices should be a function of the trading process. Market participants learn from the sequence of trades, update their beliefs, and this causes prices to move. Since the behavior of the stochastic process of prices is behind most of the questions studied in financial economics, it becomes fundamental to understand this learning mechanism.

As the literature shows, every feature of the trading process which is correlated with the value of the asset, may provide information to market participants. For example, in Easley and O'Hara (1987), the trade size is what provides information, but in Easley and O'Hara (1992) is the timing of trades. From an empirical perspective, there is no clear consensus about what actually drives the relation between trades and prices. For Jones et al. (1994) is the occurrence of transactions per se, and not their size, what contains relevant information for pricing securities. However, Huang and Masulis (1999) and Chan and Fong (2000) conclude that trade size contains no trivial information. Dufour and Engle (2000) report that both the trade duration and the trade size are informative. Finally, Kempf and Korn (1999) report a non-linear relationship between the trade size and the price impact. These conflicting findings suggest that trade size could be an unsatisfactory indicator of the information risk and that traders learn from more complex interactions of several trade features.

A related topic deals with the estimation of the theoretical components of the bid ask spread. Adverse selection costs (Bagehot, 1971) are usually characterized as the permanent impact that a trade-related shock produces on the equilibrium value of the stock. Current methods are based on structural models with an exogenous trading process characterized either by a buy sell indicator (e.g., Huang and Stoll, 1997) or by the trade size (e.g., Glosten and Harris, 1988). However, there are serious concerns about the ability of these models to measure adverse selection costs. Recently, Van Ness et al. (2001) examine the relation between adverse selection costs estimators and corporate finance indicators of information asymmetry. They conclude that structural models perform weakly. The results are similar to the ones obtained using the posted spread and therefore bring into question the added benefit of these theoretical measures. Hasbrouck (1991a,b) introduced an alternative reduced-form approach where the permanent impact of a trade can be estimated through the impulse-response function (IRF) of a vector autoregressive (VAR) model for quotes and trades. In the context of an order-driven market, de Jong et al. (1996) study the price impacts of trading using two alternative specifications: the Glosten (1994) model and the VAR model. Once again, in this context the information content of trades is characterized only by the trade size. They show that the estimates of the average adverse selection costs based on the Hasbrouck's model are twice as large of those of the structural model. The reason for the different price effect estimates is that structural models assume that prices disseminate immediately all the information content of a trade. On the contrary, the VAR model accounts for the dynamic impact of trades. We conjecture that if the quote adjustment to trade-related shocks is progressive, informed traders would try to profit from this transitory erroneous

2

pricing. This might cause persistency in the information-asymmetry risk and unusual short-term market conditions. Hence, the analysis of the quote-adjustment period after trade-related shocks becomes an attractive but almost ignored issue.

In this paper, we will analyze the learning process from the study of IBM, a NYSE-listed stock. Our main goal is to describe the dynamics of the price discovery process after a trade, contingent on its information asymmetry risk. In particular, we focus on three questions: How long does it take for market participants to learn from trades? How does this learning process depend on the expected risk of the trade? How does it affect the behavior of traders and the conditions of the market? In order to answer these questions, we will measure the expected information content of each IBM trade performed in February June 1996 using the IRF of a generalized Hasbrouck's VAR model. One of the methodological innovations will be that the information risk of a particular trade is a function of several observable trade features and simultaneously of certain market conditions, allowing for a more accurate characterization of the expected price impact. Thus, two trades of equal size could be perceived as having different information risk if they are executed under different market conditions. Another methodological improvement will be that we provide an estimation of the time (in number of events) that quotes need to incorporate all the information content of a particular trade. Huang and Stoll (1996), using a non-parametric procedure, computed the price impact of a trade over a "long enough" time horizon in order to incorporate all the information into the prices. However, this time horizon was arbitrary and constant. We want to find evidence that the periods of price adjustment depend on trade features, on market environments and on the timing of the trade.

Our findings will reinforce the relevance of the trading process as a determinant of the stochastic process of prices, although that relationship will be more complex than the one represented in a simple one-period adverse selection costs model. In summary, we will show that the market accelerates after risky trades. The trading frequency augments following trades with a high expected-risk. Indeed, the progressive quote alignments after trades originate sequences of trades with a similar (but decreasing) information-asymmetry risk. This suggests competition between informed traders (e.g., Holden and Subrahmanyam, 1992), which quickens the dissemination of the new trade-inferred information. Consequently, price discovery improves after risky trades. We will also evidence that quotes adjust faster to the trade-inferred information during the opening and closing hours of the trading session. Finally, in accordance with the short-term persistency in the information-asymmetry risk, we will report short-term anomalies in the market conditions after risky trades. Volatility and trading activity both increase and liquidity decreases with the estimated long-term impact of trades.

The structure of the paper is the following. In Section 2, we present the econometric model. In Section 3, we describe the data, discuss methodological details and report some preliminary estimation. In Section 4, we analyze the intra-daily distribution of adverse selection costs and the relative importance of this theoretical component of the spread. In Section 5, we evidence short-term persistency of the information-asymmetry risk. In Section 6, we measure the speed of adjustment of the quotes to trade-related shocks. In Section 7, we study the short-term market reaction to risky trades. Finally, in Section 8 we summarize the main empirical findings.

2. The information content of trades

The permanent impact of the unexpected component of a trade on the equilibrium value of the stock is an appropriate measure of its information-asymmetry risk. To see this, let m_t be the efficient price, the expected true value of the stock in some future end-of-trading time conditional on the public information available at moment $t (\Phi_t)$. This efficient price follows the random walk process $m_t = m_{t-1} + w_t$, where the innovation w_t is unpredictable given Φ_{t-1}. Hence, non-zero values of w_t are updates of the public information set. A shock that affects to w_t will have a permanent price impact because it alters the expected long-run value of the stock. Let x_t be a trade indicator that equals 1 for a buyer-initiated trade and -1 for a seller-initiated trade. The unexpected component of a trade is denoted by $v_{2,t}$ $(v_{2,t} = x_t - E[x_t|\Phi_{t-1}])$. Given that the predictable component of the trade is already incorporated into m_{t-1}, only $v_{2,t}$ provides new information to market participants. Hence, $E[w_t|v_{2,t}]$ is the permanent price impact of a trade. Imposing linearity, we have that $w_t = \alpha v_{2,t} + v_{1,t}$ where $v_{1,t}$ is a trade-unrelated shock $(E[v_{1,t}|v_{2,t}] = 0$ and $E[v_{1,t}v_{1,t-i}|v_{2,t}] = 0 \; \forall i \neq 0)$. Hence, $E[w_t|v_{2,t}] = \alpha v_{2,t}$. The parameter $\alpha > 0$ measures the portion of the innovation in the trading process that becomes new information. Therefore, α captures the adverse selection costs associated to the trade x_t. Structural models (e.g., Madhavan et al., 1997 and Huang and Stoll, 1997) build on similar constructions to estimate α from observable quote and trade data. However, these methods end up with an estimation of a constant average adverse selection costs for all trades in the sample, while we are looking for a procedure that allow us to estimate the expected risk of a particular trade, say α_t.

Under the hypothesis that the public information set is exclusively given by the past evolution of trades and quotes, Hasbrouck (1991a,b) introduced a reduced-form method to model the dynamic relationship between the trading process and the subsequent adjustment of market quotes. This methodology is based on a general VAR model for the quote midpoint changes and for the trade indicator x_t previously defined. Following Dufour and Engle (2000), in this paper we use a generalization of the Hasbrouck (1991a) model,

$$\Delta q_t = \sum_{i\,1}^{\infty} a_i \Delta q_{t-i} + \sum_{i\,0}^{\infty} \left[\alpha_i^q + \beta_i^{q'} \mathrm{MC}_{t-i} + \sum_{h/4} \lambda_h^q D_{t-i}^h \right] x_{t-i} + v_{1,t},$$

$$x_t = \sum_{i\,1}^{\infty} c_i \Delta q_{t-i} + \sum_{i\,1}^{\infty} \left[\alpha_i^x + \beta_i^{x'} \mathrm{MC}_{t-i} + \sum_{h/4} \lambda_h^x D_{t-i}^h \right] x_{t-i} + v_{2,t},$$

$$(1)$$

where $\Delta q_t = (q_t - q_{t-1})$ is the change in the quote midpoint after the trade x_t. The terms $v_{1,t}$ and $v_{2,t}$ are the formerly introduced zero-mean mutually and serially uncorrelated stochastic processes. We assume that the market participants learn from the trade features and the market environment. Therefore, the impact of a trade depends on a set of exogenous variables included in the vector MC_t that characterizes the trade and the market conditions. Vectors β_i^q and β_i^x have dimension $k \times 1$, where k is the number of variables in MC_t. The vector D_t contains dummies that locate the trade inside the trading session.

Microstructure theory suggests several indicators that may be correlated with the value of the asset. Easley and O'Hara (1987), among others, suggest that large-sized trades may hide impatient traders with a perishable information advantage. Hasbrouck (1991a) and de Jong et al. (1996) evidence the relevance of the trade size in the VAR methodology. Easley and O'Hara (1992) propose a model in which a reduction in the time between consecutive trades (trade duration) is an indicator of new information arriving at the market. Dufour and Engle (2000) test the predictions of Easley and O'Hara's model using the VAR methodology. Copeland and Galai (1983), and French and Roll (1986), among others, manifest the relevance of price volatility in determining liquidity in general and market quotes in particular. As far as we know, this is the first paper that incorporates volatility into the VAR framework. Finally, it is well known that adverse selection costs and liquidity are negatively related (e.g., Kyle, 1985). Following Lee et al. (1993), in this paper liquidity is measured by both immediacy costs and depth. In model (1), all these variables interact with the trade indicator to determine the long-run impact of a particular trade. Therefore, the estimated impact of a large-sized trade should depend on the quoted spread, the market depth, the price volatility, the trade duration, etc.

The VAR methodology turns out to be more flexible than the methods based on structural models. First, the trading process is not exogenous. This feature is relevant as far as a trade-related shock might cause posterior effects on the trading process (we show it does). If these dynamic effects were due to the same informative event, the initial impact would be just one part of the long-term impact of a trade-related shock. Second, if the information provided by a trade is not instantaneously incorporated into prices, the trade might also have lagged effects on quotes. The structure of the system of equations (1) accommodates all those dynamic effects on both the trading process and the market quotes. [3]

The VAR model captures, as special cases, the dynamics behind the structural models of quote formation. Indeed, the IRF of the VAR model (e.g., Sims, 1980) is an appropriate estimator of the parameter α, the long-term impact of a shock in the trading process (see Hasbrouck, 1991a). In this paper, we use the IRF of (1) as the proxy for the adverse selection costs. This IRF is conditional on the market situation and the trade features and, therefore, it is trade-specific (α_t). The Section 3 describes the database, provides the details of the derivation and implementation of the IRF, and describes the methodology used to measure the trade-specific information risk.

---

[3] The VAR model has some important econometric drawbacks. The homoskedasticity assumption in the distribution of $v_{1,t}$ and $v_{2,t}$ is restrictive given the vast evidence about intra daily deterministic patterns in volatility. Defining the model in trade time should mitigate the effect of a latent heteroskedasticity. Hasbrouck (1999) and Hausman et al. (1992) propose models for market quotes that do not assume homoskedasticity. However, in these models the trading process is endogenous. In addition, Escribano and Pascual (2000) evidence an important loss of information in averaging the quote behavior through the quote midpoint. They propose a vector error correction model for ask and bid prices, with the bid ask spread as the error correction term, that generalizes the VAR model (see also Hasbrouck, 1991a,b, 1996).

3. Data and model estimation

We use IBM trade and quote data from the TAQ database. We consider all the trading days during the first semester of 1996. IBM was one of the most frequently traded stocks during that period. This guarantees a number of observations large enough to perform the posterior empirical analysis. We only keep trades and quotes from the primary market (NYSE). Trades not codified as ''regular trades'' are not considered. These trades, out of sequence, cancelled or corrected due to errors, represent less than the 0.1% of the entire sample. Trades with the same price and time stamp are treated as just one trade. All the quotes and trades registered before the opening or after the closing of the sessions are dropped. The overnight changes in quotes are treated as missing values. Quotes with bid ask spreads lower than or equal to zero or quoted depth equal to zero are also eliminated. Price and quote files are coupled using the so-called ''5 s rule'' (see Lee and Ready, 1991). This rule assigns to each trade the first quote stamped at least 5 s before the trade itself. A trade is classified as buyer (seller) initiated when the price is closer to the ask (bid) price than to the bid (ask) price. Henceforth, the first ones are called ''buys'' and the second ones are called ''sells''. The indicator x_t equals 1 for buys, -1 for sells, and 0 for trades with execution price equal to the quote midpoint. The changes in quotes $\Delta q_t = (q_t - q_{t-1})$ are computed as the difference between the quotes that correspond to the trade x_{t+1} and to the trade x_t.

Five exogenous variables are included in the vector MC_t. The number of shares (V_t) measures the trade size. The time in seconds since the preceding trade (T_t) is the trade duration. The quoted bid ask spread measures immediacy costs (S_t). The average number of shares offered at the best ask and bid prices is the quoted depth (QD_t). [4] Volatility (R_t) is computed as the implicit volatility in the time series of $\Delta q_t(\sigma_t^2)$. It is obtained using the GARCH(1,1) model (2), estimated by maximum likelihood with the robust variance covariance matrix of Bollerslev and Wooldridge (1992). It offers the best fitting among all the models tested, including ARCH and EGARCH (e.g., Bollerslev et al., 1992). All coefficients are highly statistically significant. [5]

---

[4] An anonymous referee suggested that asymmetric depth could be more correlated with adverse selection costs than the average depth. Adverse selection costs, inventory control and barrier theories about asymmetric depth are discussed in Huang and Stoll (1994) and Engle and Patton (2000). Unfortunately, the VAR model cannot accommodate this variable. It is easy to check that, independently of the proxy used, the VAR model is not useful to determine which one of these competing theories is the appropriate theoretical framework.

[5] Although GARCH family models have been widely applied to financial time series, there are few examples of GARCH models applied to not equally spaced time series (e.g., Bollerslev and Melvin, 1994). For this reason, we have repeated the analysis with other volatility measures. They are constructed using the quote midpoint changes observed during a given time interval (from 1 to 5 min) before the time stamp of each trade. The absolute total change, the accumulated absolute and squared changes and the difference between the maximum and the minimum values of the quote midpoint during each interval are some of the measures considered. The VAR estimations are consistent across proxies and the conclusions unaltered.

$$\begin{cases} \Delta q_t = (0.0309)\varepsilon_{t-1} + \varepsilon_t \\ \sigma_t^2 = (2.2E-5) + (0.0235)\sigma_{t-1}^2 + (0.9664)\varepsilon_{t-1}^2 \end{cases} \tag{2}$$
$$\text{Adj} - R^2 : 0.001497, \quad \text{Prob}(F) = 0.0000.$$

We define eight trading-time dummy variables: one for the first half-hour of trading, five for each hour in the 10:00 15:00 interval and, finally, two for the last half-hour intervals. Therefore, we isolate the opening and closing periods of the session. All of the dummy variables, except the one corresponding to the 12:00 13:00 interval (D_t^4), were initially included in the estimation, interacting with contemporaneous and lagged values of the trade indicator x_t. Nonetheless, preliminary F tests showed that only the dummy variables affecting to the contemporaneous value of x_t were jointly statistically significant. Moreover, only the dummy corresponding to the first trading interval (9:30 10:00) became statistically significant at the 1% level. Eq. (3) shows the VAR model finally estimated.

Table 1
Estimation of the VAR model

Explanatory variables	Dependent variables		Explanatory variables	Dependent variables	
	Δq_t	x_t		Δq_t	x_t
Δq_{t-1}	**27.2**	**4174.8**	$x_t R_t$	**509.9**	
Δq_{t-2}	**22.5**	**296.8**	$x_{t-1} R_{t-1}$	18.5	**13905**
Δq_{t-3}	**24.8**	**590.8**	$x_{t-2} R_{t-2}$	98.8	**7018.6**
Δq_{t-4}	**16.3**	**383.8**	$x_{t-3} R_{t-3}$	36.6	2871.3
Δq_{t-5}	**13.7**	**383.8**	$x_{t-4} R_{t-4}$	43.2	**8484.2**
x_t	0.7408		$x_{t-5} R_{t-5}$	237.9	2427.3
x_{t-1}	**2.4121**	**301.2**	$x_t S_t$	116.8	
x_{t-2}	0.2684	**127.7**	$x_{t-1} S_{t-1}$	17.6	**853.7**
x_{t-3}	0.5438	**48.5**	$x_{t-2} S_{t-2}$	**7.25**	**102.7**
x_{t-4}	0.6255	**55.5**	$x_{t-3} S_{t-3}$	**5.72**	83.7
x_{t-5}	**1.7085**	**44.4**	$x_{t-4} S_{t-4}$	3.9	51.6
$x_t V_t$	**8.84E-04**		$x_{t-5} S_{t-5}$	**10.8**	**118.4**
$x_{t-1} V_{t-1}$	**2.99E-04**	**0.0023**	$x_t QD_t$	**0.0329**	
$x_{t-2} V_{t-2}$	**3.53E-05**	**0.0021**	$x_{t-1} QD_{t-1}$	0.0026	**0.3559**
$x_{t-3} V_{t-3}$	3.43E-05	**0.0018**	$x_{t-2} QD_{t-2}$	**0.0072**	**0.0849**
$x_{t-4} V_{t-4}$	2.72E-05	**0.0014**	$x_{t-3} QD_{t-3}$	**0.0053**	**0.1871**
$x_{t-5} V_{t-5}$	**6.23E-06**	**0.0022**	$x_{t-4} QD_{t-4}$	0.0013	0.0344
$x_t T_t$	**0.0607**		$x_{t-5} QD_{t-5}$	**0.0035**	0.0186
$x_{t-1} T_{t-1}$	**0.0189**	0.0101	$x_t D_t^1$	**1.642**	
$x_{t-2} T_{t-2}$	0.0004	**0.231**	$x_{t-1} D_{t-1}^1$		**34.2**
$x_{t-3} T_{t-3}$	0.0024	**0.265**	Adj R^2:	0.2001	
$x_{t-4} T_{t-4}$	0.0005	**0.1674**	No. obs.:	125164	
$x_{t-5} T_{t-5}$	**0.0094**	0.1246	F(42,125164):	746.58	Prob > F: 0.0000

This table reports the Generalized Least Squares (GLS) coefficient (×1000) estimates of the VAR model in (3) using all IBM trades from January to June 1996. Coefficients in **bold** indicate significant at the 1% level. V_t trade size (in number of shares). T_t time (in seconds) since the preceding trade. S_t bid ask spread. QD_t quoted depth (average between depth at the ask and depth at the bid prices). R_t volatility (implicit volatility of Δq_t estimated with a GARCH (1,1) model), q_t quote midpoint, x_t 1 for buys, 1 for sells and 0 otherwise.

$$\Delta q_t = \sum_{i\,1}^{5} a_i \Delta q_{t-1} + \sum_{i\,0}^{5} \left[\alpha_i^q + \beta_i^{q'} \mathrm{MC}_{t-i}\right] x_{t-i} + \lambda_1^q D_t^1 x_t + v_{1,t}$$

$$x_t = \sum_{i\,1}^{5} c_i \Delta q_{t-1} + \sum_{i\,1}^{5} \left[\alpha_i^x + \beta_i^{x'} \mathrm{MC}_{t-i}\right] x_{t-i} + \lambda_1^x D_t^1 x_{t-1} + v_{2,t}. \tag{3}$$

Table 1 displays the estimated coefficients of the VAR model (3) using all trades in the sample. Several tests show that ordinary least squares (OLS) residuals are heteroskedastic but not autocorrelated. Hence, we estimate the model using generalized least squares (GLS). The estimated coefficients are consistent with those found in previous studies and are consistent with the predictions of adverse selection costs models. A large-sized trade, executed a few seconds after the previous trade, within an illiquid and price-volatile period has a larger expected impact on quotes. The initial impact is statistically significant for all exogenous variables (interacting with the trade indicator). Lagged effects are especially relevant for trade-size, immediacy costs and trade durations. The trade equation shows the strong positive autocorrelation of signed trades previously evidenced by Hasbrouck (1991a). In summary, a shock in the trading process produces an instantaneous adjustment reinforced with later dynamic adjustments on both the trading activity and the market quotes. This implies progressive rather than immediate adjustments to trade-related shocks.

Finally, we approximate the vector moving average representation of (3) by Monte Carlo simulation. The IRF provides an estimation of the accumulated impact of each trade on quotes, conditional on market conditions and trade features. We call this conditional accumulated impact $I_t(\Delta q_t | v_{2,t}, \mathrm{MC}_t, D_t)$. A larger expected impact means a higher information-asymmetry risk assigned to the trade. The simulated impact for the case of infinite order polynomials appears in Eq. (4). The asterisk means a simulated value, not observed. [6]

$$\widetilde{I}_t(\Delta q_t | v_{2,t}, \mathrm{MC}_t, D_t) = \sum_{y\,1}^{n} a_y \left[\sum_{i\,1}^{n-y} \Delta q_{t+i}^*\right] + \sum_{y\,0}^{n} \alpha_y^q \left[\sum_{i\,1}^{n-y} x_{t+i}^* + v_{2,t}\right] + \sum_{y\,0}^{n} (\beta_y^q)^{\mathrm{T}}$$
$$\times \left[\sum_{i\,1}^{n-y} (\mathrm{MC}_{t+i}^* + \lambda^{q'} D_{t+i}^*) x_{t+i}^* + (\mathrm{MC}_t + \lambda^{q'} D_t) v_{2,t}\right]. \tag{4}$$

---

[6] To perform the simulation, we must define the generating process of the exogenous variables. We have assumed that each of them follows a general probabilistic process exogenous to the VAR model (3) that we have approximated by a linear autoregressive (AR) model. We include dummies that control for deterministic intra daily patterns. Furthermore, the expected impact of a trade should only depend on the impact of preceding trades. For this reason, the IRF for the first trade in February is obtained using the VAR model estimated with all trades in January (around 24.000 trades). For the second trade in February, the VAR model is estimated with all trades in January but the first one and adding the first trade in February, and so on. In this manner, the sample used to estimate (3) changes for each simulated trade but the sample size remains constant. Simultaneously, the coefficients of the VAR model are revised trade after trade. Due to space limitations, we do not provide all the details of the simulations run. However, they are included in the working paper version of this paper, which can be downloaded in pdf format from the web page http://www.eco.uc3m.es/personal/cv/alvaroe.html.

In the empirical analysis $n = 50$ and, as in Hasbrouck (1991a,b), de Jong et al. (1996), and Dufour and Engle (2000), the polynomials in the VAR model (3) are truncated at lag five. Because of the definition of x_t, midpoint trades are not simulated ($x_t = 0$). Moreover, the conditional expectation of x_t has to take values in the range of possible values $[-1, 1]$ during all the simulation steps. This may not be the case for extreme values of MC_t. These observations have been detected and dropped. After all, we estimated the IRF of nearly 80.000 trades.

To compute the adverse selection, we first locate the step τ_t of the simulation that reaches the 99% of the total estimated impact. Notice that τ_t is an estimator of the time (in number of posterior trades) required for prices to reflect all the information conveyed by the trade. The accumulated impact at this point is our estimation of the adverse selection costs of that trade (ASC_t). [7] Using the percentiles of the empirical distribution of the absolute value of ASC_t, trades are classified in five groups, from lower to higher expected risk. A trade belongs to ASC(1) if $|ASC_t| < P(0.25)$, to ASC(2) if $P(25) \leqslant |ASC_t| < P(50)$, to ASC(3) if $P(50) \leqslant |ASC_t| < P(75)$, to ASC(4) if $P(75) \leqslant |ASC_t| < P(95)$ and, finally, to ASC(5) if $|ASC_t| \geqslant P(95)$, where $P(j)$ represents the value of the j% percentile. Reference values are $P(25) = 0.0503$, $P(50) = 0.0629$, $P(75) = 0.0855$ and $P(95) = 0.1292$. The median proportion of ASC_t explained by the initial shock is 23.09%, with an interquartile range of 11.12%, and the median proportion after the first five simulation steps is 72.02% (15.05). Therefore, once all the dynamics are taken into account, it is observed that an important part of the long-run impact of a trade is associated to the posterior dynamics. This result is consistent across the five risk levels.

4. Preliminary findings

4.1. The intra-daily distribution of adverse selection costs

Wei (1992), Foster and Viswanathan (1990, 1993), and Madhavan et al. (1997) suggest that adverse selection costs are not uniformly distributed throughout the day. These costs decrease towards the end of the session, together with an increase in inventory holding costs (see Madhavan et al., 1997). This finding is consistent with having a higher concentration of information-motivated, versus liquidity-motivated, traders, during the initial intervals of the trading session. We are able to check this hypothesis using ASC_t.

Fig. 1 shows the empirical distribution of the IBM trades by trading interval and adverse selection costs level, measured by ASC_t. We divide the session in thirteen half-hour intervals. Bands of the same color represent the percentage of trades belonging to ASC(j), $j = \{1, \ldots, 5\}$, executed in each interval (the 13 bands of the

---

[7] We have also considered the 50%, 75% and 90% of the total estimated impact to define ASC_t. Spearman rank correlations are significantly superior to the 95%. Neither the classification of the simulated trades nor the empirical findings in the next sections are remarkably affected by the percentage considered.
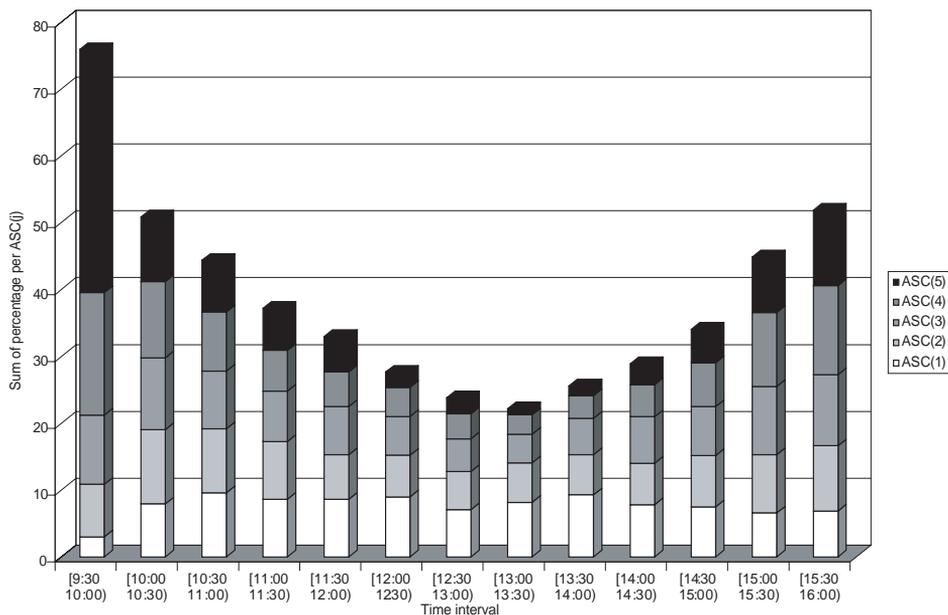
Fig. 1. Intra daily distribution of the information asymmetry risk. Figure shows the percentage of trades in each ASC(j) category, j 1, ..., 5, that were performed in each half hour interval of the trading session.

same color sum to the 100%). The column height is the sum of all five percentages per interval. The distribution of the trading activity exhibits the usual U-shaped pattern. The trades with the highest expected adverse selection costs, ASC(4) and ASC(5), concentrate at the edges of the session. The 47.62% of all trades belonging to ASC(5) were performed during the opening (36.37%) and closing (11.25%) half-hours. Similarly, the 31.66% of all trades classified as ASC(4) were accomplished during these intervals, 18.32% only during the first half-hour. In contrast, ASC(1) trades are detected mainly in the middle of the session and only the 2.99% in the first half-hour of trading. Previous results manifest that the risk of trading with an informed agent is the highest during the opening interval. The ASC(5) and ASC(4) trades represent more than the 50% of all trades observed between 9:30 and 10:00 a.m. At the closing of the sessions, these trades are the 32.15% of all trades executed. On the contrary, ASC(1) and ASC(2) trades represent the 25.14% of trades during the opening period versus the 67.62% between 1:00 and 1:30 p.m. and the 64.45% between 1:30 2:00 p.m. In the next sections, we evidence that this non-uniform distribution of informed trading is reflected in the price discovery speed.

4.2. The relevance of adverse selection costs in a dynamic context

Using data from the Paris Bourse, de Jong et al. (1996) estimate the adverse selection costs component of the bid ask spread using two alternative approaches, a

structural model based on Glosten (1994) and the Hasbrouck's VAR model. These authors assume that only the trade size provides information to the market participants and, hence, the permanent price impact of trades depends only on this variable. They show that the estimates of the average adverse selection costs based on the Hasbrouck's model are twice as large of those of the structural model. The reason is that the Glosten's model assumes that prices immediately disseminate all the information conveyed by a trade and therefore ignores the lagged price effects. Indeed, this is a usual feature of adverse selection costs models (e.g., Glosten and Harris, 1988; Madhavan et al., 1997 and Huang and Stoll, 1997). In this section, we extend the de Jong et al.'s analysis in several ways: we use a more complete characterization of the price impact (ASC_t), we compare the results for trades with different risk levels, we control for the intra-daily regularities observed in the previous subsection and, finally, we use NYSE data.

We use Lin et al. (1995) methodology as the theoretical referent. [8] Eq. (5) summarizes this method,

$$(q_t - q_{t-1}) = \delta(P_t - q_{t-1}) + e_t, \tag{5}$$

where P_t represents the execution price of the trade, e_t is the error term, $|P_t - q_{t-1}|$ is the half-effective spread and the parameter δ measures adverse selection costs. Notice that δ is the percentage of the effective spread that is not realized due to the immediate quote-change after the trade. Under the assumption that trades incorporate at once the information content of the trade, this immediate change equals the total price impact. With the sample of simulated trades, we estimate Eq. (5) by OLS robust to heteroskedasticity and autocorrelation of unknown form (Newey and West, 1987). We obtain that the average adverse selection costs represent the $\hat{\delta} = 26.3\%$ of the effective spread.

The first column of coefficients in Table 2 shows the results of estimating Eq. (6), which is a generalization of (5) where we control for the trading interval and the risk of asymmetric information. The D_t^h, $h = \{1, \ldots, 8\}$, are dummy variables to control for the trading interval. The Q_t^j, $j = \{1, \ldots, 5\}$, are the dummy variables representing the risk of asymmetric information, that is Q_t^j equals 1 if $ASC_t \in ASC(j)$ and 0 otherwise. The variable u_t is the error term.

$$\Delta q_t = \delta_0(P_t - q_{t-1}) + \sum_{j\;2}^{5}[\delta_Q^j(P_t - q_{t-1})]Q_t^j + \sum_{\substack{h\;1 \\ h/4}}^{8}[\delta_D^h(P_t - q_{t-1})]D_t^h + u_t. \tag{6}$$

Table 2 shows that the percentage of total immediacy costs due to adverse selection costs increases significantly with $ASC(j)$, running from the 15.97% for $ASC(1)$ trades to the 29.95% for $ASC(5)$ trades. Moreover, the trades executed during the

---

[8] We choose Lin et al. (1995) because it is one of the most used models in practice (e.g., Brockman and Chung, 1999). It does not require the estimation of dynamic equations; then, the results will be unaffected by the trades we removed (see Section 3). Finally, its parametric simplicity facilitates generalizations. Van Ness et al. (2001) compare several structural models and conclude that no single model appears to perform better than the others.

Table 2
Adverse selection costs over the total immediacy costs

Coefficient ($\times 100$)[a]	Δq_t	Initial impact (simulation)
δ_0	**15.9793**	**16.1386**
δ_Q^2	**5.1832**	**2.8813**
δ_Q^3	**10.5893**	**5.1355**
δ_Q^4	**12.2063**	**6.4202**
δ_Q^5	**13.9802**	**10.1475**
δ_D^1 [9:30 10:00]	**2.5499**	**1.9521**
δ_D^2 [10:00 11:00]	1.2797	**1.0046**
δ_D^3 [11:00 12:00]	0.7079	0.2678
δ_D^5 [13:00 14:00]	0.5919	0.0478
δ_D^6 [14:00 15:00]	0.6355	0.0242
δ_D^7 [15:00 15:30]	1.2583	**0.7867**
δ_D^8 [15:30 16:00]	0.6813	**0.7202**
Adj R^2 (NW)	0.2055	0.2121

This table summarizes the results of estimating the percentage of the effective spread that is due to adverse selection costs. The Lin et al.'s (1995) model has been extended in order to control for intra daily effects and the risk level due to information asymmetries, see Eq. (6). The model is estimated by OLS with the Newey and West (1987) robust method. Two alternative dependent variables have been used: the observed change in the midpoint of the bid ask spread (the original variable in Lin et al., 1995) and the initial (first step) impact estimated by the simulation of the VAR model in (3).
[a] Format in bold means significant at the 1% level.

first half-hour of trading have a 2.55% risk premium. The second column of coefficients in Table 2 contains the estimation of (6) but replacing the dependent variable Δq_t with the initial impact of the trade obtained by the simulation of the VAR model (3). The percentages are similar to those obtained with the observed data but, in this case, the trades accomplished during the final interval of the session have also a risk premium. Therefore, if only the initial impact of the trade is considered, adverse selection costs represent, in average terms, no more than the 30 32% of the effective spread. However, this conclusion changes when the dynamic effects of trades are taken into consideration. As in de Jong et al. (1996), we compute the ratio of the corresponding ASC_t value to the half-effective spread, $\widetilde{I}_t^\tau(\Delta q_t | v_{2,t}, MC_t, D_t)/(P_t - q_{t-1})$. The median simulated total price impact represents the 80% for ASC(3) trades, the 90% for ASC(4) trades and more than the 100% for ASC(5) trades. This result is consistent with de Jong et al. (1996) finding that large trades of the Paris Bourse have a permanent price impact larger than the quoted bid ask spread. This finding manifests that adverse selection costs in a dynamic framework are far more important than the one-period structural models would suggest. The quoted spread for a frequently traded stock (most of the time equal to the tick, US\$ 1/8 in 1996) may not compensate for the costs of providing liquidity to trades with a high risk. This result reflects two reasonable issues. First, informed traders prefer to trade when the stock is liquid. Second, the specialist duty of maintaining stable liquidity conditions forces him/her to offer spreads that could be insufficient to compensate high-risk levels. These losses, however, would be compensated with the liquidity-motivated traders.

5. Risk persistency

Previous sections have shown the relevance of the dynamic impacts of trade-related shocks on both the quotes and the trading process. The information content of a trade is not instantaneously incorporated into prices, suggesting that market participants take some intervals of trading to have their expectational differences resolved (e.g., Harris and Raviv, 1993). Under these circumstances, it would be reasonable to find additional trades taking profits from the temporal divergence between market quotes and the efficient price. That is, we should observe sequences of trades with similar (but decreasing) values of ASC_t In this section we study the expected short-run risk persistency by modeling the time series of ASC_t.

The usual unit-root tests (extended Dickey and Fuller, 1979, and Phillips and Perron, 1988) show that the ASC_t time series is a $I(0)$ process. Moreover, the autocorrelation and partial autocorrelation functions indicate that ASC_t can be modeled as an AR process of finite order $(AR(p))$, with p at least equal to 3. Both the information inferred from the trading process and the possible transitory deviation between the efficient price and q_t are expected to increase with $ASC(j)$, $j = \{1, \ldots, 5\}$. Thus, our intuition is that the magnitude of the AR coefficients of the $AR(p)$ model should also increase with the estimated ASC_t. We proceed with the estimation of the truncated AR(5) model, see Eq. (7), for the time series of ASC_t using the GLS method. The dummies Q_t^j, $j = \{1, \ldots, 5\}$, were defined in Section 4.2. These dummy variables consider five thresholds in the AR structure of ASC_t. The u_t is the error term of the model. Table 3 summarizes the estimation results.

Table 3
Risk persistency

Coefficient[a]	Q_t^1: ASC(1)	Q_t^2: ASC(2)	Q_t^3: ASC(3)	Q_t^4: ASC(4)	Q_t^5: ASC(5)
ϕ_1^j	**0.3333**	**0.4082**	**0.5896**	**0.7427**	**0.8964**
	(0.0118)	(0.0193)	(0.0120)	(0.0189)	(0.0470)
ϕ_2^j	**0.1279**	**0.1229**	**0.1191**	**0.0755**	0.0285
	(0.0142)	(0.0186)	(0.0131)	(0.0175)	(0.0690)
ϕ_3^j	**0.1222**	**0.0907**	**0.0436**	0.0257	0.0719
	(0.0140)	(0.0184)	(0.0136)	(0.0174)	(0.0794)
ϕ_4^j	**0.1007**	**0.0918**	**0.0729**	**0.0421**	0.0153
	(0.0138)	(0.0114)	(0.0109)	(0.0155)	(0.0461)
ϕ_5^j	**0.1513**	**0.1693**	**0.1298**	**0.1402**	**0.1291**
	(0.0114)	(0.0119)	(0.0097)	(0.0158)	(0.0469)

Adj R^2 0.9497, Prob $> F$ 0.0000.
This table shows the estimated GLS coefficients of the truncated AR(5) model in (7). Robust standard errors are in parenthesis. The time series ASC_t is built with the estimated adverse selection costs corresponding to all IBM trades executed from February to June 1996. A trade belongs to the set ASC(1) if $|ASC_t| < P(0.25)$, to ASC(2) if $P(25) \leqslant |ASC_t| < P(50)$, to ASC(3) if $P(50) \leqslant |ASC_t| < P(75)$, to ASC(4) if $P(75) \leqslant |ASC_t| < P(95)$ and, finally, to ASC(5) if $|ASC_t| \geqslant P(95)$, where $P(y)$ represents the value of the $y\%$ percentile of the empirical distribution of ASC_t. The dummy Q_t^j equals 1 if $ASC_t \in ASC(j)$, j $\{1, \ldots, 5\}$, and 0 otherwise.
[a] Format in bold means statistically significant at the 1% level.

$$\text{ASC}_t = \sum_{j=1}^{5} \left(\sum_{p=1}^{5} \phi_p^j \text{ASC}_{t-p} \right) Q_t^j + u_t. \tag{7}$$

Table 3 reveals significant differences in the autocorrelation structure of the time series ASC_t across the five levels of adverse selection costs. Using the Wald test (e.g., Davidson and MacKinnon, 1993) we reject (at the 1% level) the null hypothesis that the sums of the AR coefficients corresponding to each pair $\text{ASC}(j)$ and $\text{ASC}(k)$, with $j \neq k$, are equal. Indeed, the sum of the AR coefficients increases with $\text{ASC}(j)$, $j = \{1, \ldots, 5\}$. This means that trades with high information-asymmetry risk (large ASC_t value) are likely followed by similar trades more than those with low information-asymmetry risk (low ASC_t value). These clusters of risky trades suggest that there is competition between informed traders. Informed traders try to maximize their gains exploiting the temporal disagreement between the quoted prices and the efficient price. Table 3 reveals that, after the execution of a risky trade, the information-asymmetry risk persists due to the gradual adjustment of market quotes. This short-term risk persistency increases with the adverse selection costs associated to the trade at time t. [9]

We have also considered an alternative specification of Eq. (7) that uses the trade-time dummy variables D_t instead of Q_t in order to truncate the AR(5) structure. This new specification would capture differences in the AR(5) coefficients per trading hour. However, the statistical tests performed do not reject the null hypothesis of an equal AR(5) structure across trading hours. Therefore, we conclude that the results of Table 3 are due to differences in adverse selection costs and are not biased by intra-daily regular patterns.

6. Price discovery and the information content of trades

Besides the ASC_t measure, the simulation procedure of the VAR model (3) produces an additional output. This output is the time (in number of events) required by quotes to disseminate all the information content in a particular trade, say τ_t. We have transformed it into real time using the distance in seconds between the time stamp of the simulated trade and the time stamp of the τ_t^{th} trade afterwards. We denote $D(\tau)_t$ the time series formed by the real-time distances for all trades. In this section, we evaluate how long does it take for market participants to learn from trades by studying whether the time of adjustment depends on the market conditions and the characteristics of trades (MC_t). That is, we check how does the price discovery process depend on each trade's expected risk. Table 4 summarizes the results of estimating Eq. (8) by OLS with the Newey and West (1987) robust standard errors.

---

[9] Risk persistency could also be evaluated by applying the extended Dickey Fuller unit roots test to each threshold in (7). However, the t statistics of such a test are neither standard nor currently tabulated. It should be necessary to obtain the critic values by simulation, something that is out of the scope of this paper.

Table 4
The speed of learning from the information content of trades

Variable	Coefficient[a]	Std. Error
δ_0 (const.)	**721.41**	7.987
V_t (δ_1)	**0.0051**	0.00018
T_t (δ_2)	**2.1679**	0.05745
R_t (δ_3)	**26495.11**	3001.8
S_t (δ_4)	**424.72**	22.827
QD_t (δ_5)	**0.4391**	0.02076
D_t^1 [9:30 10:00]	**263.98**	6.5971
D_t^2 [10:00 11:00]	**232.71**	6.4680
D_t^3 [11:00 12:00]	**108.26**	7.2188
D_t^5 [13:00 14:00]	7.6262	8.6188
D_t^6 [14:00 15:00]	**123.29**	7.3578
D_t^7 [15:00 15:30]	**225.56**	7.0505
D_t^8 [15:30 16:00]	**276.85**	6.7039
Adj. R^2	0.2346	
Prob $> F$	0.0000	

This table summarizes the estimation of Eq. (8) by OLS robust to heteroskedasticity and autocorrelation (Newey and West, 1987). The variable τ_t is an estimation of the time (in number of events) that quotes need to capture all the information provided by a given trade. This τ_t comes from the simulation of the VAR model (3) and is trade specific. $D(\tau)_t$ is the series formed with all τ_t expressed in real time (s). V_t trade size (in number of shares). T_t time (in s) since the preceding trade. S_t bid ask spread. QD_t quoted depth (average between depth at the ask and depth at the bid prices). R_t volatility (implicit volatility of Δq_t estimated with a GARCH (1,1) model), and D_t^j, j {1,...,8}, are dummy variables that control for deterministic intra daily patterns.

[a] Format in bold means statistically significant at the 1% level.

$$D(\tau)_t = \delta_0 + \delta_1' MC_t + \sum_{\substack{j/4}}^{8} \gamma_j D_t^j + u_t. \tag{8}$$

For simplicity, we assume linearity in the specification. In order to control for the regular patterns in trading frequency, we include the dummy variables D_t^h (with 12:00 13:00 as the control interval).

As expected, the duration of the learning process depends on the moment of execution. During the less frequently traded hours (between 12:00 and 14:00), the period of quote adjustment could go on around 12 min ($\delta_0/60$). However, if the trade is executed during the first half-hour of the trading session, this time is reduced to 7.5 min (($\delta_0 + \gamma_1$)/60), approximately. Moreover, Table 4 reveals that the adjustment period is reduced with trade size and the volatility of prices. On the contrary, the adjustment period is increased with liquidity and trade duration. Collectively, the higher the expected adverse selection costs, the shorter the adjustment period. If we replace MC_t in Eq. (8) by the ASC_t estimator, we end up with the same conclusion.

Our conjecture is that this finding is due to an increase in the trading intensity following trades with high information-asymmetry risk. This effect accelerates the process of price discovery. The increase in the trading intensity may reflect the sequential reaction of the market to the same informative signal, an imitative behavior of other

agents in the market, or even order-splitting by the same agent (see Easley and O'Hara, 1987; Biais et al., 1995, and He and Wang, 1995). From the results in the previous section we suggest that the temporal disagreement between quoted prices and the efficient price may induce competition between informed traders. Admati and Pfleiderer (1988), Easley and O'Hara (1992), and Holden and Subrahmanyam (1992) develop alternative models in which competition between informed traders favors price efficiency, especially if their activity is based on the same informative signal. Consistently, our results imply that prices respond more quickly after a trade if the estimated risk increases. Additionally, price discovery is faster during periods of risky trading concentration. In Section 7, we provide further evidence supporting our hypothesis of "market acceleration" after an informative trade.

Huang and Stoll (1996) measured the impact of a trade at time t by $(q_{t+\tau^*} - q_t)x_t$, where $q_{t+\tau^*}$ is the quote midpoint associated to the first trade executed (at least) τ^* minutes later. The value of τ^* is the same for all trades and arbitrarily fixed. Huang and Stoll use this measure to compare the adverse selection costs of a matched sample of NYSE and Nasdaq listed stocks. The evidence in this section indicates that the results of Huang and Stoll may be biased because the value of τ^* depends on the moment of execution, the concrete characteristics of the trade, and the market conditions. Moreover, the value of τ_t for a given trade might differ under different microstructures. The difference between our measure of adverse selection costs and the effective spread could be seen as an ex-ante and more flexible version of the Huang and Stoll's realized spread.

7. The short-term market response to risky trades

In this section, we study the market impact of both the progressive adjustment of market quotes and the associated risk persistency reported in previous sections. We analyze how the market behavior after a trade depends on its expected information-asymmetry risk. There is a large literature about unusual market patterns around localized informative events (e.g., Lee et al., 1993; Koski and Michaely, 2000, and Goldstein and Kavajecz, 2000). Unusual patterns generally consist on increases in trading activity and volatility, and reductions in liquidity both before and after the event. Pre-event behavior is attributed to informed traders that anticipate the informative shock. Post-event behavior is more difficult to interpret. If the public disclosure resolves the information asymmetry, the market should return to its pre-event behavior. Kim and Verrecchia (1994) develop a model in which certain traders are able to make superior judgments from public disclosures than others. This situation increases the information asymmetry after the event, reduces liquidity and increases the trading activity and the volatility (see also Harris and Raviv, 1993).

Event studies compare the periods surrounding the events under analysis with a benchmark that is not influenced by such (or other) informative events. Such a methodology is not workable in our case because our events (trades) are not isolated from other similar events. We have reported short-term persistency in the information-asymmetry risk caused by clusters of trades that can be differentiated by its average

level of adverse selection costs. We understand that clusters of trades with a similar ASC(j) level can be associated to the same event. Hence, to avoid possible biases in posterior tests, we proceed by filtering the sample. When we observe a sequence of buys or sells that are very close to each other and with the same ASC(j) level, we only include the first trade of the sequence in the subsequent tests. [10] Furthermore, results in Table 4 indicated that the impact of a trade takes on average around 12 min to be negligible. Hence, we will consider the 15 min interval going after the execution of each trade. We focus on the post-event period because the adverse selection costs estimator (ASC$_t$) measures the permanent impact of the unanticipated component of the trade.

For each minute $m = \{1, \ldots, 15\}$ we compute the following variables: (a) the number of shares traded (Vol$_{t+m}$) and (b) the number of trades completed (NT$_{t+m}$). These two variables measure trading activity. (c) The standard deviation of the quote midpoint (VQ$_{t+m}$) measures volatility. (d) The average bid ask spread (SPT$_{t+m}$) and (e) the average quoted depth (DPT$_{t+m}$), weighted by time, which stand for liquidity. For trades time stamped during the last quarter-hour of the trading session, these variables are treated as missing for the minutes that include or exceed the official closing hour (16:00 h). For each minute and using the filtered sample, we estimate Eq. (9) by GLS. The dummy variables D_t^h and Q_t^j were defined in Section 4.2. The variable S_i^m is also one of the market indicators previously discussed. Our hypothesis is that, after a trade, trading activity and price volatility increase and liquidity decreases with the trade's expected information content.

$$S_i^m = \sum_{h=1}^{8} \sum_{j=1}^{5} \alpha_{h,j} D_i^h Q_i^j + \varepsilon_i^m. \tag{9}$$

We have shown that adverse selection costs are not uniformly distributed throughout the trading session. Moreover, activity, liquidity and volatility indicators also show intra-daily regular patterns (e.g., Jain and Joh, 1988, and McInish and Wood, 1992). In Eq. (9), trades differ by the corresponding adverse selection costs level and by the moment of execution. Thus, we test for differences in market behavior after trades accomplished during the same hourly interval. Accordingly, S_i^m is standardized by trading interval. [11] Likelihood ratio tests (not reported) were used to compare the model (9) with the alternative specification in which the Q_t^j dummies

---

[10] We compute the median (in s) between two consecutive trades belonging to the same ASC(j) level, with $j \in \{1, \ldots, 5\}$. These medians are: 30 for ASC(1) trades, 24 for ASC(2) and ASC(3) trades, 17 for ASC(4) trades and 13 for ASC(5) trades. If the time between two consecutive trades of the same type is less than the corresponding median, these trades are considered as originated by the same informative event. The analysis has been repeated using other filters and even using all trades in the sample. The main findings are consistent.

[11] The standardization method is robust to outliers. For example, consider the observation that corresponds to the accumulated volume during the fifth minute after a trade time stamped at 9:58:00 (Vol$_{10:02:00\ 10:02:59}$). To standardize it we subtract the median of Vol for all the minutes traded in the period 10:00 11:00 during all the sample period. This difference is divided by the corresponding interquartile range.
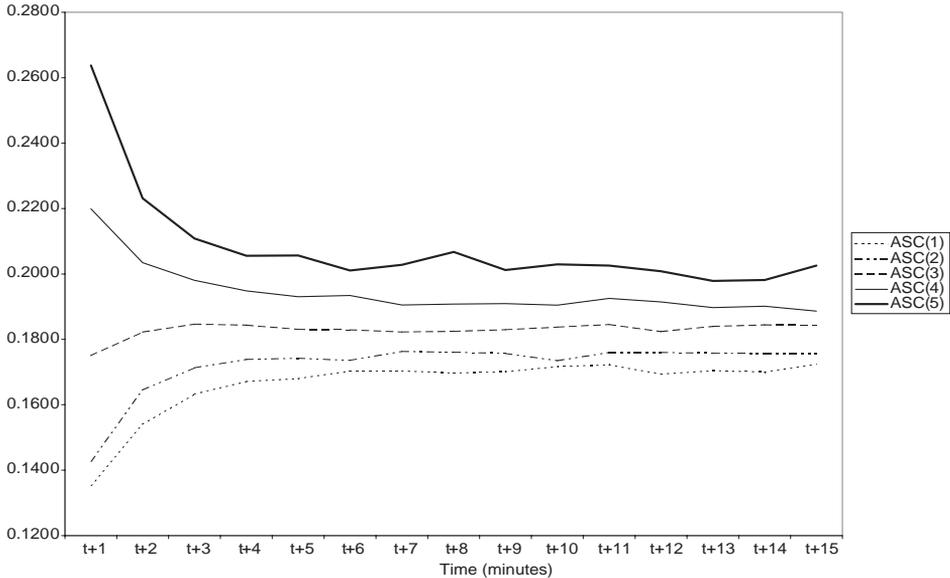
Fig. 2. Bid ask spread dynamics after a trade conditional on adverse selection costs. Figure shows the average bid ask spread weighted by time (not standardized) for each information asymmetry risk level ($ASC(j)$, j $\{1, \ldots, 5\}$) during the 15 min after a trade.

were removed. We rejected the null hypothesis of equality of the two specifications for all S_i^m and for all intervals. This implies that the expected information-asymmetry risk of trades provide information about the posterior market behavior, although the value of the test statistic slowly decreases as we move away from the initial impact.

The main findings are summarized as follows. [12] As expected, immediacy costs are increasing with adverse selection costs. For some trading intervals, these differences persist during the 15 min analyzed. Fig. 2 shows the average bid ask spread weighted by time (not standardized) for each $ASC(j)$ level, $j = \{1, \ldots, 5\}$, during the 15 min interval. Differences are not statistically significant 6 or 7 min after the trade. We obtain similar statistically significant differences in volatility; it increases with $ASC(j)$, implying that trades with greater adverse selection costs tend to increase the uncertainty about the true value of the stock. With respect to the quoted depth, the $ASC(5)$ ($ASC(1)$) trades are located in periods of higher (lower) depth than the other trades. However, there is no monotonic increasing relationship between the quoted depth and $ASC(j)$. After the trade, the quoted depth tends to increase with less intensity (or even decrease) as the trade risk increases. Results are not a clear cut.

We find an unusually intense trading activity going after risky trades. Supporting our "market acceleration" hypothesis, trades with a high expected risk lead to an in-

---

[12] Due to space limitations, we do not provide the tables with the estimated coefficients of model (9). However, they are included in the working paper version of this paper, which can be downloaded from http://www.eco.uc3m.es/personal/cv/alvaroe.html.

crease in trading intensity, probably due to the successive reaction of the market to the new information and to the competition between traders. Hence, the wider spreads may be the result of the combination of a "liquidity consumption effect" caused by the increase in trading activity and a greater protectionism by liquidity providers facing a greater risk of informed trading. In an attempt to judge the relevance of the consumption effect in explaining unusual spreads and volatility, we also estimate Eq. (10) with $S_i^m = \{\text{SPT}_i^m, \text{VQ}_i^m\}$ standardized.

$$S_i^m = \text{Vol}_i^m + \text{NT}_i^m + \sum_{h\,1}^{8} \sum_{j\,1}^{5} \alpha_{h,j} D_i^h Q_i^j + \varepsilon_i^m. \tag{10}$$

We obtain that the differences in immediacy costs are not completely explained by a liquidity consumption effect. Indeed, the bid ask spread and the volatility, once corrected by trading activity, still increase with the adverse selection costs level $(\text{ASC}(j), j = \{1, \ldots, 5\})$.

Globally, these findings suggest that market participants learn from the observable features of the trade and the characteristics of the market. They revise their positions altering (at least) the liquidity of the stock and the intensity of trading. These unusual market conditions are maintained due to the short-term persistency of the information-asymmetry risk. It is important to remark that this behavior is independent of the moment of the session.

8. Conclusions

In this paper, we have described the learning process of market participants from trading and we have studied how do they incorporate the trade-related information into the stochastic process of prices. We have focused on the price discovery process of IBM, a NYSE-listed stock. Our main concern was to evaluate the dynamics of the price discovery process after a trade as a function of its information-asymmetry risk. Hence, we have evaluated the risk of each IBM trade by estimating its permanent price impact. The price impact depends on the simultaneous consideration of several aspects of the trade and the market. This represents a more refined characterization of adverse selection costs since in previous studies they typically identified the information content of trades with their size. Our estimator is trade-specific and is based on the IRF of a VAR model. This reduced-form approach accurately measures the total impact of a trade by taking into account the dynamic effects of trades on both the market quotes and on the trading process. Furthermore, we have also estimated the time that quotes take to disseminate all the information content of a particular trade.

Our empirical analysis has shown that the price discovery process accelerates after risky trades. The trading frequency increases following trades with a high expected informational content. Additionally, the progressive alignments of the quotes after a trade originate clusters of trades with a similar (but decreasing) information-asymmetry risk and, consequently, short-term risk persistency. This causes short-term

anomalies in market conditions that augment with the trade risk. Therefore, our findings strengthen the relevance of the trading process in explaining the stochastic process of prices, but also draw a scenario that is far more complex than suggested by one-period adverse selection costs models.

The analysis of the quote-adjustment period after trade-related shocks becomes an attractive but almost unexplored issue in microstructure research. The findings of this paper suggest several lines for further research, including the comparison of the price discovery process of matched samples of stocks listed in markets with different microstructures and the study of the differences in the learning process between frequently and infrequently traded stocks.

Acknowledgements

References

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. The Review of Financial Studies 1, 3 40.

Bagehot, W., 1971. The only game in town. Financial Analysts Journal 27 (12 14), 22.

Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. The Journal of Finance 50, 1655 1689.

Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modelling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5 59.

Bollerslev, T., Melvin, M., 1994. Bid ask spreads and volatility in the foreign exchange market. An empirical analysis. Journal of International Economics 36, 355 372.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. Econometric Reviews 11, 143 172.

Brockman, P., Chung, D.Y., 1999. Bid ask spread components in an order driven environment. The Journal of Financial Research 22, 227 246.

Chan, K., Fong, W. M., 2000. Trade size, order imbalance, and the volatility volume relation. Journal of Financial Economics 57, 247 273.

Copeland, T.E., Galai, D., 1983. Information effects on the bid ask spread. The Journal of Finance 38, 1457 1469.

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press.

de Jong, F., Nijman, T., Roell, A., 1996. Price effects of trading and components of the bid ask spread on the Paris Bourse. Journal of Empirical Finance 3, 193 213.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427 431.

Dufour, A., Engle, R.F., 2000. Time and the price impact of a trade. Journal of Finance 55 (6), 2467 2498.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69 90.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. The Journal of Finance 47 (2), 577 605.

Engle, R.F., Patton, A.J., 2000. Impacts of trades in a error correction model of quote prices. Working Paper, University of California, San Diego.

Escribano, A., Pascual, R., 2000. Dynamic asymmetries in bid ask responses to the innovations in the trading process. Working Paper #00 47 (18), Universidad Carlos III de Madrid.

Foster, F.D., Viswanathan, S., 1990. A theory of the intraday variations in volume, variance and trading costs in securities markets. The Review of Financial Studies 3, 593 624.

Foster, F.D., Viswanathan, S., 1993. Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models. The Journal of Finance 48, 187 211.

French, K.R., Roll, R., 1986. Stock return variances. The arrival of information and the reaction of traders. Journal of Financial Economics 17, 5 26.

Glosten, L.R., 1994. Is the electronic open limit order book inevitable? Journal of Finance 49, 1127 1161.

Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123 142.

Goldstein, M.A., Kavajecz, K.A., 2000. Liquidity provision during circuit breakers and extreme market conditions. Working Paper #2000 02, NYSE.

Harris, M., Raviv, A., 1993. Differences of opinion make a horse race. The Review of Financial Studies 6, 473 506.

Hasbrouck, J., 1991a. Measuring the information content of stock trades. The Journal of Finance 46, 179 207.

Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis. The Review of Financial Studies 4, 571 595.

Hasbrouck, J., 1996. Modeling market microstructure time series. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 14, Statistical Methods in Finance. Elsevier, North Holland, Amsterdam.

Hasbrouck, J., 1999. The dynamics of discrete bid and ask quotes. Journal of Finance 54 (6), 2109 2142.

Hausman, J.A., Lo, A.W., MacKinlay, A.C., 1992. An ordered probit analysis of transaction costs prices. Journal of Financial Economics 31, 319 379.

He, H., Wang, J., 1995. Differential information and dynamic behavior of stock trading volume. Review of Financial Studies 8, 919 972.

Holden, C.W., Subrahmanyam, A., 1992. Long lived private information and imperfect competition. The Journal of Finance 47, 247 270.

Huang, R.D., Masulis, R.W., 1999. Informed trading activity and stock price volatility: Evidence from the London Stock Exchange. Working Paper 99 09, Owen Graduate School of Management, Vanderbilt University.

Huang, R.D., Stoll, H.R., 1994. Market microstructure and stock return predictions. Review of Financial Studies 7 (1), 179 213.

Huang, R.D., Stoll, H.R., 1996. Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics. 41, 313 357.

Huang, R.D., Stoll, H.R., 1997. The components of the bid ask spread: A general approach. The Review of Financial Studies 10, 995 1034.

Jain, P., Joh, G. H., 1988. The dependence between hourly prices and trading volume. Journal of Financial and Quantitative Analysis 23, 269 284.

Jones, C.M., Kaul, G., Lipson, M.L., 1994. Transactions, volume, and volatility. Review of Financial Studies 7 (4), 631 651.

Kempf, A., Korn, O., 1999. Market depth and order size. Journal of Financial Markets 2, 29 48.

Kim, O., Verrecchia, R.E., 1994. Market liquidity and volume around earnings announcements. Journal of Accounting and Economics 17, 41 67.

Koski, J.L., Michaely, R., 2000. Prices, liquidity, and the information content of trades. Review of Financial Studies 13, 659 696.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315 1335.

Lee, C.M., Mucklow, B., Ready, M.J., 1993. Spreads, depths, and the impact of earnings information: An intraday analysis. The Review of Financial Studies 6, 345 374.

Lee, C.M., Ready, M.J., 1991. Inferring trade direction from intraday data. The Journal of Finance 46, 733–746.

Lin, J.-C.C., Sanger, G.C., Booth, G.G., 1995. Trade size and components of the bid–ask spread. Review of Financial Studies 8, 1153–1183.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? A transaction-level analysis of NYSE stocks. The Review of Financial Studies 10, 1035–1064.

McInish, T.H., Wood, R.A., 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. The Journal of Finance 47, 753–764.

Newey, W.K., West, K.D., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Sims, C., 1980. Macroeconomics and reality. Econometrica 48, 1–49.

Van Ness, B.F., Van Ness, R.A., Warr, R.S., 2001. How well do adverse selection components measure adverse selection? Financial Management 30, 77–98.

Wei, P.H., 1992. Intraday variations in trading activity, price variability and the bid–ask spread. Journal of Financial Research 15, 265–276.