Adverse Selection Costs, Trading Activity
and Price Discovery in the NYSE: An Empirical Analysis.*

Roberto Pascual
Departamento de Economía y Empresa
Universidad de las Islas Baleares, SPAIN
E-mail: rpascual@uib.es
Phone: +34-971-17-2845
Fax: +34-971-17-3426

Alvaro Escribano
Departamento de Economía
Universidad Carlos III de Madrid, SPAIN
E-mail: alvaroe@eco.uc3m.es
Phone: +34-91-624-9854
Fax: +34-91-624-9329

Mikel Tapia
Departamento de Economía de la Empresa
Universidad Carlos III de Madrid, SPAIN
E-mail: mtapia@emp.uc3m.es
Phone: +34-91-624-9639
Fax: +34-91-624-9608

This version
(January 2002)

* This paper has benefited from the support of Spanish DGICYT project #PB98-0030 and European Union grant TMR ERB-4061 PL 97-0994. Corresponding author: Alvaro Escribano, Departamento de Economía, Universidad Carlos III de Madrid, C/Madrid, 126, Getafe, Madrid, SPAIN, E-mail: alvaroe@eco.uc3m.es, Phone: +34-91-624-9854, Fax: +34-91-624-9329.
Adverse Selection Costs, Trading Activity
and Price Discovery in the NYSE: An Empirical Analysis.

ABSTRACT
This paper studies the role the 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 impact on market quotes. The price impact depends on observable trade features and market conditions. We show that price discovery is quicker after risky trades and 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 following risky trades.

Keywords: Microstructure, adverse selection costs, trade-related information, high-frequency data.

Jel: G1
1. Introduction

A central notion of microstructure literature is that trades are informative. Accordingly, theoretical research characterizes the stochastic process of prices as a function of the trading process. Market participants learn from the sequence of trades, updating their beliefs, and this causes prices to move. Since the stochastic process of prices underlies most of the topics studied in financial economics, it becomes fundamental to understand this learning mechanism. In this paper, we analyze the learning process of a NYSE listed stock, that is, we measure the information content of its trades and evaluate how it is reflected in the market quotes. Our main concern is to describe the dynamics of the price discovery process after a trade conditional on its information asymmetry risk. We study how long market participants take to learn from trades, when the price discovery is faster, how it depends on the expected risk of the trade and how it affects the traders behavior and the market conditions.

Every feature of the trading process correlated with the value of the asset may provide information to market participants. For example, in Easley and O'Hara (1987) is the trade size that provides information, but in Easley and O'Hara (1992) is the timing of trades. From an empirical perspective, there is no consensus about what actually drives the relation between the trading activity and the process of price formation. Thus, Jones et al. (1994) conclude that the occurrence of transactions per se, and not their size, contains all the information relevant to the pricing of securities. However, Huang and Masulis (1999) and Chan and Fong (2000) conclude that trade size contains no trivial information. Similarly, Dufour and Engle (2000) report that both the trade duration and the trade size are informative. These contradictory findings suggest that traders do learn from the complex interaction of several trade features.

A closely related research is the estimation of the theoretical components of the bid-ask spread. The 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 usually build on structural models with an exogenous trading process characterized just by the trade sign (e.g., Huang and Stoll, 1997) or the trade size (e.g., Glosten and Harris, 1988). A reduced form approach is introduced by Hasbrouck (1991a,b) 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 effects of trading using two alternative approaches, the Glosten (1994) model and the VAR model. Once again, 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 effects estimates is that structural models assume that prices immediately disseminate all the information conveyed by a trade. On the contrary, the VAR model accounts for the dynamic impact of trades. If the quote adjustment to trade-related shocks is progressive, informed traders would try to profit from the transitory erroneous pricing. This might cause persistency in the information-asymmetry risk and unusual short-term market conditions.

In this paper, we measure the expected information content of each IBM trade performed at the NYSE during the first semester of 1996 using the IRF of a generalization of the Hasbrouck’s VAR model. A methodological innovation is that the information risk of a particular trade is a function of several observable trade features and market conditions acting simultaneously, allowing for a more accurate characterization of the price impact than in previous studies. Our approach takes into account the dynamic impacts on both the market quotes and the trading process, and we will evidence that these effects represent a non-negligible part of the total impact of a trade. Huang and Stoll (1996) also address the trade-specific estimation of adverse selection costs. Using a non-parametric procedure, they compute the price impact over a time horizon “long enough” as to allow all the information conveyed by the trade to be incorporated into prices. However, this time horizon is arbitrary and independent of the trade features. On the contrary, our procedure provides an estimation of the time (in number of events) that quotes need to take in all the information content of a particular trade. We evidence that this period depends on the trade features, the market environment and the moment of the session.

We use these instruments to analyze how IBM traders learn from the trading process. Our main empirical findings are the following. First, the market accelerates after risky trades. We show that trading frequency augments following trades with a high expected risk, perhaps due to competition between informed traders (e.g., Admati and Pfleiderer, 1988, and Holden and Subrahmanyan, 1992). Indeed, quote alignments after trades are progressive rather than instantaneous and this fact originates sequences of trades with a similar (but decreasing) information-asymmetry risk. This apparent competition quickens the dissemination of the new trade-inferred information. Consequently, price discovery improves after risky trades. Second, it is evidenced that quotes adjust faster to the new information inferred from trades during the opening and closing hours of the trading session. Third, in accordance with the short-term persistency in the information-asymmetry risk, we observe 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 paper proceeds as follows. Section 2 presents the econometric model. Section 3 describes the data, discusses some methodological details and shows preliminary estimation results. Section 4 analyzes the intra-daily distribution of adverse selection costs and measures the relative importance of the adverse selection costs component of the bid-ask spread. Section 5 evidences short-term persistency in the information-asymmetry risk after trades. Section 6 examines the speed of adjustment of market quotes to a trade-related shock. Section 7 studies the short-term market reaction to risky trades. Finally, section 8 concludes.

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, understood as the expected true value of the stock in some future end-of-trading time conditional on the public information available at moment \( t (Φ_t) \). Let this efficient price follow the random walk process \( m_t = m_{t-1} + w_t \), where the innovation \( w_t \) is unpredictable given all public information. Hence, non-zero values of \( w_t \) should be understood as updates of the public information set. A shock that affects to \( w_t \) will have a permanent 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 \( v_{2,t} \) (i.e., \( v_{2,t} = x_t - E[x_t | F_{t-1}] \)). 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, the permanent price impact of a trade will be given by \( E[w_t | v_{2,t}] \).

Imposing linearity, we have that \( w_t = αv_{2,t} + v_{1,t} \), where \( v_{1,t} \) is a trade-unrelated shock (\( E[v_{1,t} | v_{2,t}] = 0 \) and \( E[v_{1,t}, v_{1,i} | v_{2,t}] = 0 \) for \( i ≠ 0 \)). Hence, \( E[w_t | v_{2,t}] = αv_{2,t} \). In this model the parameter \( α > 0 \) measures the portion of the innovation in the trading process \( (v_{2,t}) \) that becomes new information and that actualizes the expectation about the true value of the stock \( (m_t) \). Therefore, \( α \) captures the adverse selection costs associated to the trade \( x_t \). Microstructure structural models (e.g., Madhavan et al., 1997, and Huang and Stoll, 1997) build on similar structures to estimate the parameter \( α \) from observable quote and trade data. However, these methods end up with an estimation of the average adverse selection costs for all trades in the sample. For our purposes, we require a procedure to estimate the information-asymmetry risk of each 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) introduces an econometric methodology 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 changes of the quote midpoint and the trade indicator $x_t$ previously defined. Following Dufour and Engle (2000), in this paper we use a generalization of the Hasbrouck (1991a) model given by,

$$
\Delta q_t = \sum_{i=1}^{\infty} a_i \Delta q_{t-i} + \sum_{i=0}^{\infty} \left[ \alpha^q_i + \beta^q_i MC_{t-i} + \sum_{h \neq 4} \lambda^h_i D_{t-i} \right] x_{t-i} + v_{1,t} \\

x_t = \sum_{i=1}^{\infty} c_i \Delta q_{t-i} + \sum_{i=0}^{\infty} \left[ \alpha^c_i + \beta^c_i MC_{t-i} + \sum_{h \neq 4} \lambda^h_i D_{t-i} \right] x_{t-i} + v_{2,t}.
$$

The revision in market quotes $\Delta q_t = (q_t - q_{t-1})$ represents the change in the quote midpoint after a trade in period $t$ ($x_t$). 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 their 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. 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 less 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. We introduce volatility in the VAR model as a determinant of the trade risk. Finally, it is 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. All these variables interact with the trade indicator to determine the long-run impact of a particular trade. Thus, the estimated impact of a large-sized trade will depend on the quoted spread, the market depth, the price volatility, the trade duration, and so on. Vectors $\beta^q_i$ and $\beta^c_i$ have dimension $k \times 1$, where $k$ is the number of variables in $MC_t$. The vector $D_t$ contains dummy variables that locate the trade inside the trading session.

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 (we show it does) on the trading process. 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 (we show it is not), the trade might also
have lagged effects on quotes (we show it has). The structure in (1) accommodates these
dynamic effects on both the trading process and the market quotes.\(^1\)

The VAR model captures, as special cases, the main dynamics behind the structural
models of quote formation. Indeed, the impulse-response function (IRF) of the VAR model
(e.g., Sims, 1980) is an appropriate estimator of the parameter \( \alpha \), that is, an accurate measure
of the long-term impact of an unexpected shock in the trading process (see Hasbrouck,
1991a). In this paper, we use the IRF associated to the VAR model (1) as a proxy for the
adverse selection costs associated to each trade \( (\alpha_t) \). This IRF is conditional on the market
situation and the trade features and, therefore, trade-specific. The next section 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-asymmetry risk.

3. Data and model estimation

The VAR model (1) is estimated for IBM using trade and quote data from the TAQ
database of the NYSE. All the trading days from January to June 1996 are considered. IBM
was one of the most frequently traded stocks during this 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
discarded. These are trades cancelled or corrected due to error and trades out of sequence.
Non-regular trades 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 quote and trade registers before the
opening or after the close 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
have also been eliminated. Price and quote files are coupled using the so-called “five seconds
rule” (see Lee and Ready, 1991). This rule assigns to each trade the first quote stamped at
least five seconds before the trade itself.

Following previous empirical studies, 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 trade indicator $x_t$ equals 1 for buys, -1 for sells, and 0 for trades with execution price equal to the quote midpoint. A change in quotes is computed as the difference between the quotes that correspond to the trade at $t+1$ and to the trade at $t$, $\Delta q_t = (q_t - q_{t-1})$. Eight trading-time dummies are constructed: one for the first half-hour of trading, five for each trading hour between 10:00 a.m. and 3:00 p.m. and, finally, the last trading hour has been divided in two half-hour intervals. This procedure isolates the opening and closing periods of the session.

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. Immediacy costs is measured by the quoted bid-ask spread ($S_t$). Quoted depth ($QD_t$) is computed as the average between the number of shares offered at the best ask and bid prices. Volatility ($R_t$) is measured by 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 and with the robust variance-covariance matrix of Bollerslev and Wooldridge (1992) (see Bollerslev et al., 1992 for a review of these models). It offers the best fitting among all the models tested, including ARCH and EGARCH. All coefficients are highly statistically significant.

$$\begin{align*}
\Delta q_t &= (0.0309)e_{t-1} + \varepsilon_t \\
\sigma_t^2 &= (2.2E-5) + (0.0235)\sigma_{t-1}^2 + (0.9664)e_{t-1}^2 \\
Adj - R^2 &= 0.001497, \text{ Prob}(F) = 0.0000
\end{align*}$$

Equation (3) shows the VAR model finally estimated.

$$\begin{align*}
\Delta q_t &= \sum_{i=1}^{5} c_{i} \Delta q_{t-i} + \sum_{i=0}^{5} \left[ \alpha_i^q + \beta_i^q MC_{t-i} \right] x_t + \lambda_i^q D_t^1 x_t + v_{1t} \\
x_t &= \sum_{i=1}^{5} c_{i} \Delta q_{t-i} + \sum_{i=0}^{5} \left[ \alpha_i^q + \beta_i^q MC_{t-i} \right] x_t + \lambda_i^q D_t^1 x_{t-i} + v_{2t}.
\end{align*}$$

Table I displays the estimated coefficients of the VAR model (3) using all trades executed from January to June 1996 (both included). Several tests show that Ordinary Least Squares (OLS) residuals are heteroskedastic but not autocorrelated. Hence, we estimate the model
using Generalized Least Squares (GLS). Results coincide with those found by other researchers (see Hasbrouck, 1991a; Dufour and Engle, 2000) 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 already evidenced by Hasbrouck (1991a). In conclusion, a shock in the trading process will produce an instantaneous but also posterior dynamic adjustments on both the trading activity and the market quotes. This implies progressive rather than immediate adjustments to trade-related shocks.

The VMA representation of (3) is approximated by Monte Carlo simulation because the VAR model is non-linear due to the vector of exogenous variables. The IRF provides an estimation of the accumulated impact of each trade on quotes, conditional on market conditions and trade features. Denote this conditional accumulated impact by $I_t(\Delta q_t, v_{2,t}, MC_t, D_t)$. A larger expected impact means a higher information-asymmetry risk assigned to the trade. In the Appendix A, we detail the steps of the simulation exercise.

The procedure described in Appendix A depends on a preliminary estimation of model (3). The expected impact of a trade should depend on the impact of similar preceding trades but not on the impact of posterior trades. For this reason, the IRF (A.3) for the first trade in February is obtained using the VAR model (3) estimated with all trades in January (around 24,000 trades). For the second trade in February, the VAR model is estimated with the data of all the 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. At the same time, the coefficients of the VAR model are revised trade after trade. This procedure implies that trades in January cannot be simulated. Midpoint trades in the February-June sample are not simulated either. Because of the definition of the trade indicator $x_t$, it is not possible to simulate trades with execution price equal to the quote midpoint ($x_t = 0$). Finally, 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 $MC_t$. Through the simulation, these observations are detected and discarded. At the end, we have estimated the IRF of nearly 80,000 trades.
To compute the adverse selection costs that correspond to a given trade ($ASC_t$), we first locate the step $\tau_t$ of the simulation that reaches the 99% of the total estimated impact. Notice that the variable $\tau_t$ will be an estimator of the time (in number of trades) required for prices to reflect all the information carried by the trade. The accumulated impact at this point is our estimation of the adverse selection costs for a particular trade. We have also considered the 50%, 75% and 90% of the total estimated impact to define $ASC_t$. Spearman Rank Correlations among the time series obtained in each case 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. Using the percentiles of the empirical distribution of the absolute value of $ASC_t$, trades are classified in five groups, $ASC(1)$ to $ASC(5)$, from lower to higher expected risk. A trade belongs to $ASC(1)$ if $|ASC_t| < P(0.25)$, to $ASC(2)$ if $P(25) \leq |ASC_t| < P(50)$, to $ASC(3)$ if $P(50) \leq |ASC_t| < P(75)$, to $ASC(4)$ if $P(75) \leq |ASC_t| < P(95)$ and, finally, to $ASC(5)$ if $|ASC_t| \geq P(95)$, where $P(y)$ represents the value of the $y$% 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 provoked by the initial trade 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. This costs decrease towards the end of the session, together with an increase in inventory holding costs (see Madhavan et al., 1997). This finding would be consistent with a higher concentration of information-motivated versus liquidity-motivated traders during the initial intervals of the trading session. We check this hypothesis using $ASC_t$.

[Figure 1]

Figure 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 $ASC(j), j=\{1,...,5\}$,
executed in each interval (the thirteen bands of the 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), are concentrated at the extremities 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 the opening and closing half-hours, and the 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 of each session. The ASC(5) and ASC(4) trades represent more than the 50% of all trades observed between 9:30 a.m. and 10:00 a.m. At closing, 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 p.m. and 1:30 p.m. and 64.45% between 1:30 p.m. and 2:00 p.m. In the next sections, we evidence that this non-uniform distribution of informed trading is reflected in the speed of the price discovery.

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 feature. 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 explanation is that the Glosten’s model assumes that prices immediately disseminate all the information conveyed by a trade and, therefore, it ignores the lagged price effects. Indeed, this is a common 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 this analysis by using a more complete characterization of the price impact (ASC_t), comparing the results for trades with different risk levels, controlling for the intra-daily regularities shown in previous subsection and using NYSE data.

We use Lin et al. (1995) methodology as the theoretical point of reference.\footnote{Equation (4) summarizes this method,
\[
(q_t - q_{t-1}) = \delta(P_t - q_{t-1}) + e_t,
\]
where \(P_t\) represents the execution price of the trade, \(e_t\) is the error term and \(|P_t - q_{t-1}|\) is the half-effective spread. The parameter \(\delta\) measures adverse selection costs. Notice that \(\delta\) 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 equation (4) by OLS robust to heteroskedasticity and autocorrelation of unknown form (Newey and West, 1987). We obtain that the average adverse selection costs represent the \(\delta = 26.3\%\) of the effective spread.

The first column of coefficients in Table II shows the result of estimating equation (5), a generalization of (4) where we control for the trading interval and the risk of asymmetric information. The variables \(D_t^h\), \(h = 1, \ldots, 8\), are the dummies for the trading interval. The variables \(Q_t^j\), \(j = 1, \ldots, 5\), are the dummies for the risk of asymmetric information, where \(Q_t^j\) equals 1 if \(ASC_t \in ASC(j)\) and 0 otherwise. The variable \(u_t\) is the error term. We use again the Newey-West method.

\[
\Delta q_t = \delta_0 (P_t - q_{t-1}) + \sum_{j=2}^{5} \left[ \delta_j (P_t - q_{t-1}) \right] Q_t^j + \sum_{h=1}^{8} \left[ \delta_h^h (P_t - q_{t-1}) \right] D_t^h + u_t
\]

[Table II]

Table II shows that the percentage of total immediacy costs due to adverse selection costs significantly augments with \(ASC(j)\), from the 15.97\% for \(ASC(1)\) trades to the 29.95\% for \(ASC(5)\) trades. Moreover, the trades executed during the first half-hour of trading have a 2.55\% risk premium. The second column of coefficients in Table II contains the estimation of (5) but replacing the dependent variable \(\Delta q_t\) with the initial impact of the trade obtained by the simulation of the VAR model (3). The percentages are much the same as those obtained with the observed data but, in this case, the trades accomplished during the final interval of the session also have a risk premium. We can conclude that, 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, the conclusion changes when the dynamic effects of trades are taken into consideration. As in de Jong et al. (1996), we compute the absolute ratio of the corresponding \(ASC_t\) value to the half-effective spread, \(\left| T_t^- (\Delta q_t | V_{2,t}, MC_t, D_t ) | (P_t - q_{t-1}) \right|\).
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 the costs of providing liquidity to trades with a high risk. This result reflects two very reasonable facts. 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. Therefore, 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. We should expect sequences of trades with similar (but decreasing) values of \( ASC_t \). This section studies this expected short-run persistency in the information-asymmetry risk 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 autoregressive 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 \( ASC_j \), \( j \in \{1, ..., 5\} \). Thus, our intuition is that the magnitude of the autoregressive 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 (6) for the time series of \( ASC_t \) using the Generalized Least Squares (GLS) method. The dummies \( Q_{jt} \), \( j \in \{1, ..., 5\} \), were introduced in section 4.2. These dummy variables define five thresholds in the autoregressive structure of \( ASC_t \). The variable \( u_t \) is the error term of the model. Table III summarizes the estimation.
Table III reveals significant differences in the autocorrelation structure of the time series \( ASC_t \) across the five levels of adverse selection costs considered. Using the Wald test (e.g., Davidson and MacKinnon, 1993) we reject (at the 1% level) the null hypothesis that the sums of the autoregressive coefficients corresponding to each pair \( ASC(j) \) and \( ASC(k) \), with \( j \neq k \), are equal. Indeed, the sum of the autoregressive coefficients increases with \( ASC(j) \), \( j = \{1,...,5\} \). This means that trades with an elevated information-asymmetry risk (large \( ASC_t \) value) are more likely followed by similar trades than those with low information-asymmetry risk (low \( ASC_t \) value). These clusters of trades with similar risk suggest competition between informed traders. Informed traders try to maximize their gains exploiting the temporal disagreement between the quoted prices and the efficient price. Table III reveals that the information-asymmetry risk persists after the execution of a risky trade because of the gradual adjustment of market quotes. This short-term risk persistency does augments with the adverse selection costs associated to the trade at time \( t \).

We have considered an alternative specification for (6) that uses the trade-time dummy variables \( D_t \) instead of the \( Q_t \) ones to truncate the AR(5) structure. This specification would capture differences in the AR(5) coefficients per trading hour. 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 in Table III do are due to differences in adverse selection costs and are not biased by any intra-daily regular pattern.

6. Price discovery and the information content of trades

The simulation procedure of the VAR model (3) produces an additional output to the trade-specific adverse selection costs measure \( (ASC_t) \). This output is an estimation of the time (in number of events) that quotes require to disseminate all the information content in a particular trade, say \( \tau_t \). 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 \( \tau \)th trade afterwards. We denote by \( D(\tau_t) \) to the series formed by the real-time distances obtained for all trades. In this section, we evaluate how long market participants take to learn from trades by studying whether this time of adjustment depends on the market conditions and the characteristics of trades \( (MC_t) \).
That is, we check how the price discovery process depends on each trade’s expected risk. Table IV summarizes the results of estimating equation (7) by Ordinary Least Squares (OLS) with the Newey and West (1987) robust standard errors.

$$D(\tau_i) = \delta_0 + \delta_1'MC_i + \sum_{j=1}^{s} \gamma_j D^{h}_j + u_i.$$  \hspace{1cm} (7)

For simplicity, we assume linearity in the specification. In order to control for the regular patterns in trading frequency, we include the dummy variables $D^{h}_j$ (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 minutes ($\delta_0/60$). However, if the trade is executed during the first half-hour of the trading session, this time is reduced to 7.5 minutes (($\delta_0 + \gamma_1)/60$), approximately. Moreover, Table IV reveals that the adjustment period shortens with trade size and price volatility. On the contrary, it lengthens with liquidity and the trade duration. Collectively, the higher the expected adverse selection costs, the shorter the adjustment period. If we replace $MC_i$ in equation (7) by the $ASC_i$ estimator, we end up with the same conclusion.

We hypothesize that this last 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 by 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). As the results in the previous section suggested, 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 when its estimated risk increases. Additionally, price discovery is faster during the periods where the risky trading concentrates. The next section provides additional evidence supporting our
hypothsis 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) $\tau^*$ minutes later. The value of $\tau^*$ 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 $\tau^*$ depends on the moment of execution, the concrete characteristics of the trade, and the market conditions. Moreover, the value of $\tau_t$ for a given trade might differ under different microstructures. The difference between our adverse selection costs measure 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

This section studies 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 following a trade depends on its expected information-asymmetry risk. There is by now 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 increased trading activity, volatility and illiquidity 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 non-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 and produces less liquidity and the possibility of more trading activity and 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 one to the others and with the same ASC(j) level, only the first trade of the sequence is included in the subsequent tests. Furthermore, results in Table IV indicated that the impact of a trade takes around 12 minutes, on average, to be negligible. Hence, we consider the fifteen minutes 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 \((DPT_{t+m})\) both weighted by time stand for liquidity. For trades time stamped during the last quarter-hour of the trading session (15:45-16:00 h.), these variables are treated as missing for the minutes that include or exceed the official closing hour (16:00 h.). For each minute \( m=\{1,\ldots,15\} \), we estimate equation (8) for the filtered sample by GLS. The dummy variables \( D_h^j \), \( h=\{1,\ldots,8\} \) and \( Q_j^i \), \( j=\{1,\ldots,5\} \) have been defined in section 4.2. The variable \( S_m^i \) is one of the market indicators previously discussed. Our null hypothesis is that, after a trade, trading activity and price volatility increase and liquidity decreases with the trade’s expected information content.

\[
S_m^i = \sum_{h=1}^{8} \sum_{j=1}^{5} \alpha_{h,j} D_h^j Q_j^i + e_m^i. \tag{8}
\]

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 equation (8), 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, \( S_m^i \) is standardized by trading interval. Likelihood ratio tests (not reported) are used to compare the model (8) with an alternative specification in which the \( Q_j^i \), \( j=\{1,\ldots,5\} \) dummies are removed. We reject the null hypothesis of equality of the two specifications for all \( S_m^i \) and for all intervals. This implies that the expected information-asymmetry risk of trades do provide information about the posterior market behavior, though the value of the test statistic decreases as we move away from the initial impact.
The main findings are shown in the Appendix B. Only the estimated coefficients for the opening and closing half-hours are reported. The results for the remaining intervals are very consistent and available upon request. Two types of tests are performed. First, asterisked values indicate statistically significant differences between the coefficients, in the sense that $\alpha_{h,k} > \alpha_{h,k-1}$ for $k=\{2,\ldots,5\}$, that is, $S_{i}^{m}$ increases from one risk level to the next. Second, the bold format means that the coefficient is not statistically different from zero. Table B.I evidences that immediacy costs are increasing in adverse selection costs. For some trading intervals, these differences persist during the 15 minutes analyzed. Figure 2 shows the average bid-ask spread weighted by time (not standardized) for each ASC(j) level, $j=\{1,\ldots,5\}$, during the 15 minutes interval. Differences fade away 6 or 7 minutes after the trade. Table B.II shows similar statistically significant differences in volatility. It increases with ASC(j), implying that trades with greater adverse selection costs tend to increment the uncertainty about the true value of the stock. Regarding the quoted depth (not reported), the ASC(5) (ASC(1)) trades are located in periods of higher (lower) depth than the other trades. However, there is not a 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 augments. Results, in any case, are not conclusive.

[Figure 2]

Tables B.III-B.IV report an unusually intense trading activity going after risky trades. Supporting our “market acceleration” hypothesis, trades with a high expected risk lead to an increase 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 quoted 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 an increase in the risk of informed trading. In an attempt to judge the relevance of the liquidity consumption effect in explaining unusual spreads and volatility, we have also estimated equation (9), with $S_{i}^{m} = \{SPT_{i}^{m},Q^{m}\}$ standardized.

$$S_{i}^{m} = Vol_{i}^{m} + NT_{i}^{m} + \sum_{h=1}^{5} \sum_{j=1}^{5} \alpha_{h,j} D_{i}^{h} Q_{j}^{i} + \varepsilon_{i}^{m}. \quad (9)$$

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 (ASC(j), $j=\{1,\ldots,5\}$). Results
(available upon requests) are similar to those detailed in Tables B.I and B.II.

Globally, these findings suggest that market participants learn from the observable features of the trade and the market. Then, they revise their positions altering (at least) the liquidity of the stock and the intensity of trading. These unusual market conditions keep on 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

This paper has described how market participants learn from the trading process and incorporate the trade-related information into the stochastic process of prices. We focused on the price discovery process of IBM, a NYSE-listed stock. Our main concern has been to evaluate the dynamics of the price discovery process after a trade conditional on its information-asymmetry risk. To do that, we evaluated the risk of each IBM trade in the sample by estimating its expected permanent price impact. A remarkable methodological innovation is that the price impact depends on the simultaneous incidence of several characteristics of the trade and the market. This allowed for a more precise characterization of adverse selection costs than in previous studies that typically identify the information content of trades with their size. The estimator is based on the impulse-response function of a VAR model estimated with all trades performed one month before the corresponding trade. Therefore, it is trade-specific. Furthermore, as pointed out by Hasbrouck (1996) and de Jong et al. (1996), the reduced form approach of the VAR model accurately characterizes the total impact of a trade by taking into account the posterior effects of trades on both the market quotes and the trading process itself. Finally, a further output is a trade-specific estimation of the time that quotes take to disseminate the information content of each trade-related shock.

Our empirical findings show that the price discovery process accelerates after risky trades. Trading frequency augments following trades with an expected high informational content. Presumably, this is due to the competition between informed traders. Additionally, the progressive alignment of the quotes after a trade originates clusters of trades with a similar (but decreasing) information-asymmetry risk and, consequently, short-term risk persistency. At the same time, this causes short-term anomalies in market conditions that augment with the trade risk. We also report that quotes adjust faster during the opening and closing hours of the session, where the risky trading concentrates.
Footnotes

1. The VAR model has also some important drawbacks from an econometric point of view. The homokedasticity assumption in the distribution of $v_{1,t}$ and $v_{2,t}$ seems to be restrictive given the vast evidence about intra-daily deterministic patterns in volatility. In any case, defining the model in trade-time should mitigate the effect of a latent heteroskedasticity. Hasbrouck (1999) introduces a model for ask and bid quotes in which the innovations in the efficient price are generated by an EGARCH model. Hausman et al. (1992) analyze changes in prices using an ordered probit model, without forcing homoskedasticity. However, in these models the trading process is endogenous. In addition, Escribano and Pascual (2000) show that there is an important loss of information in averaging the quote behavior through the quote midpoint. These authors propose a vector error correction model (VEC) for ask and bid prices, with the bid-ask spread as the error correction term, that generalizes the VAR model. Hasbrouck (1991a,b), Hasbrouck (1993), de Jong et al. (1996), Hasbrouck (1996), and Dufour and Engle (2000), among others, discuss other controversial aspects related to the estimation of the VAR model.

2. An anonymous referee suggested that asymmetric depth could be more correlated with adverse selection costs than the average depth. Unfortunately, the VAR model cannot accommodate this variable because: (1) The trading process is summarized using a unique trade indicator ($x_t$). (2) All exogenous variables enter into the model interacting with $x_t$. (3) All theories predict that asymmetric depth will have opposite effects on buyer-initiated and seller-initiated trades. Adverse selection costs, inventory control and barrier theories about asymmetric depth are discussed in Huang and Stoll (1994) and Engle and Patton (2000). 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. We would need a model with two indicators, one for buyer initiated and another for seller initiated trades to incorporate asymmetric depth into the analysis (for example, see Escribano and Pascual, 2000).

3. 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. These alternative volatility proxies are constructed using the quote midpoint changes observed during a given time interval (from 1 to 5 minutes) 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. In general, the VAR estimations are consistent across proxies and the main conclusions unaltered.

4. We choose Lin et al. (1995) because it is one of the most used models in practice (e.g., Brockman and Chung, 1999). Moreover, this method does not require the estimation of dynamic equations; so, our estimation results will be less affected by the trades removed (see section 3). Finally, the parametric simplicity of the Lin et al. (1995) facilitates extensions directed to control for several factors. Although Huang and Stoll (1997) is probably the most general model, its main advantage is to distinguish between inventory holding and operative costs. Our interest, however, is focused on adverse selection costs only.

5. Risk persistency could also be evaluated by applying the extended Dickey-Fuller unit roots test to each threshold in (6). 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.

6. We compute the median (in seconds) between two consecutive trades belonging to the same ASC(j) level, with $j=\{1,...,5\}$. 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 to 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.

7. 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 $\{Vol_{100200-100259}\}$. To standardize it we subtract the median of $Vol$ for all the minutes traded in the period 10:00 a.m. to 11:00 a.m. during all the sample period. This difference is divided by the corresponding interquartile range.

19
References


TABLE I
Estimation of the VAR model

This table reports the Generalized Least Squares (GLS) coefficients of the VAR model in (3) using all IBM trades from January to June 1996. \( V_t \) = trade size (in number of shares), \( T_t \) = time (in seconds) since the preceding trade. \( S_t \) = bid-ask spread. \( QD_t \) = quote depth (average between depth at the ask and depth at the bid prices). \( R_t \) = volatility (implicit volatility of \( \Delta q_t \) estimated with a GARCH(1,1) model), \( q_t \) = quote midpoint, \( x_t = 1 \) for buys, \(-1 \) for sells and \( 0 \) otherwise.

<table>
<thead>
<tr>
<th>( \Delta q_{t-1} ) (Coef. x 1000)</th>
<th>( x_t )</th>
<th>( \Delta q_{t-2} ) (Coef. x 1000)</th>
<th>( x_t )</th>
<th>( \Delta q_{t-3} ) (Coef. x 1000)</th>
<th>( x_t )</th>
<th>( \Delta q_{t-4} ) (Coef. x 1000)</th>
<th>( x_t )</th>
<th>( \Delta q_{t-5} ) (Coef. x 1000)</th>
<th>( x_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-27.2)</td>
<td>(-4174.8)</td>
<td>(22.5)</td>
<td>(296.8)</td>
<td>(24.8)</td>
<td>(590.8)</td>
<td>(16.3)</td>
<td>(383.8)</td>
<td>(13.7)</td>
<td>(383.8)</td>
</tr>
<tr>
<td>(x_{t-1}R_{t-1})</td>
<td>(-18.5)</td>
<td>(x_{t-2}R_{t-2})</td>
<td>(98.8)</td>
<td>(x_{t-3}R_{t-3})</td>
<td>(-36.6)</td>
<td>(x_{t-4}R_{t-4})</td>
<td>(-43.2)</td>
<td>(x_{t-5}R_{t-5})</td>
<td>(237.9)</td>
</tr>
<tr>
<td>(2.4121)</td>
<td>(301.2)</td>
<td>(0.2684)</td>
<td>(127.7)</td>
<td>(0.5438)</td>
<td>(48.5)</td>
<td>(-0.6255)</td>
<td>(55.5)</td>
<td>(-1.7085)</td>
<td>(44.4)</td>
</tr>
<tr>
<td>(x_{t-1}S_{t-1})</td>
<td>(17.6)</td>
<td>(x_{t-2}S_{t-2})</td>
<td>(-7.25)</td>
<td>(x_{t-3}S_{t-3})</td>
<td>(-5.72)</td>
<td>(x_{t-4}S_{t-4})</td>
<td>(-3.9)</td>
<td>(x_{t-5}S_{t-5})</td>
<td>(-10.8)</td>
</tr>
<tr>
<td>(0.000884)</td>
<td></td>
<td>(x_{t-1}V_{t-1})</td>
<td>(-0.0329)</td>
<td>(0.000299)</td>
<td>(0.0023)</td>
<td>(x_{t-2}V_{t-2})</td>
<td>(-0.0026)</td>
<td>(0.000353)</td>
<td>(-0.0021)</td>
</tr>
<tr>
<td>(x_{t-3}V_{t-3})</td>
<td>(-0.000343)</td>
<td>(x_{t-2}V_{t-2})</td>
<td>(0.0072)</td>
<td>(x_{t-3}V_{t-4})</td>
<td>(0.0053)</td>
<td>(x_{t-4}V_{t-4})</td>
<td>(0.0013)</td>
<td>(x_{t-5}V_{t-5})</td>
<td>(0.0035)</td>
</tr>
<tr>
<td>(x_{t-5}QD_{t-5})</td>
<td>(-0.0186)</td>
<td>(x_{t-1}QD_{t-1})</td>
<td>(0.3559)</td>
<td>(x_{t-2}QD_{t-2})</td>
<td>(0.0849)</td>
<td>(x_{t-3}QD_{t-3})</td>
<td>(0.1871)</td>
<td>(x_{t-4}QD_{t-4})</td>
<td>(0.0344)</td>
</tr>
<tr>
<td>(-0.0607)</td>
<td></td>
<td>(x_{t-1}D_{t-1}^i)</td>
<td>(-1.642)</td>
<td></td>
<td></td>
<td>(x_{t-2}D_{t-2}^i)</td>
<td>(34.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-0.0189)</td>
<td>(0.0101)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x_{t-3}D_{t-3}^i)</td>
<td>(-0.265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0004)</td>
<td>(-0.231)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x_{t-4}D_{t-4}^i)</td>
<td>(-0.1674)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0094)</td>
<td>(-0.1246)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(x_{t-5}D_{t-5}^i)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Format in bold means significant at the 1% level.
TABLE II
Adverse selection costs over the total immediacy costs

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 equation (5). The model is estimated by Ordinary Least Squares (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).

<table>
<thead>
<tr>
<th>Coefficient (x100)*</th>
<th>Δq_t</th>
<th>Initial impact (simulation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ_0</td>
<td>15.9793</td>
<td>16.1386</td>
</tr>
<tr>
<td>δ_1</td>
<td>5.1832</td>
<td>2.8813</td>
</tr>
<tr>
<td>δ_2</td>
<td>10.5893</td>
<td>5.1355</td>
</tr>
<tr>
<td>δ_3</td>
<td>12.2063</td>
<td>6.4202</td>
</tr>
<tr>
<td>δ_4</td>
<td>13.9802</td>
<td>10.1475</td>
</tr>
<tr>
<td>δ_0 [9:30 10:00)</td>
<td>2.5499</td>
<td>1.9521</td>
</tr>
<tr>
<td>δ_0 [10:00 11:00)</td>
<td>1.2797</td>
<td>1.0046</td>
</tr>
<tr>
<td>δ_0 [11:00 12:00)</td>
<td>0.7079</td>
<td>0.2678</td>
</tr>
<tr>
<td>δ_0 [13:00 14:00)</td>
<td>0.5919</td>
<td>0.0478</td>
</tr>
<tr>
<td>δ_0 [14:00 15:00)</td>
<td>0.6355</td>
<td>0.0242</td>
</tr>
<tr>
<td>δ_0 [15:00 15:30)</td>
<td>1.2583</td>
<td>0.7867</td>
</tr>
<tr>
<td>δ_0 [15:30 16:00)</td>
<td>0.6813</td>
<td>0.7202</td>
</tr>
</tbody>
</table>

| Adj. -R^2(NW): | 0.2055 | 0.2121 |

* Format in bold means significant at the 1% level.
TABLE III
Risk persistency

This table shows the estimated Generalized Least Squares (GLS) coefficients of the truncated AR(5) model in (6). 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 $|\text{ASC}_t| < P(0.25)$, to ASC(2) if $P(25) \leq |\text{ASC}_t| < P(50)$, to ASC(3) if $P(50) \leq |\text{ASC}_t| < P(75)$, to ASC(4) if $P(75) \leq |\text{ASC}_t| < P(95)$ and, finally, to ASC(5) if $|\text{ASC}_t| \geq P(95)$, where $P(y)$ represents the value of the y% percentile of the empirical distribution of ASC. The dummy $Q_j^t$ equals 1 if ASC$_t \in$ ASC(j), $j={1,\ldots,5}$, and 0 otherwise.

<table>
<thead>
<tr>
<th>Coefficient$^*$</th>
<th>$Q_1^t$: ASC(1)</th>
<th>$Q_2^t$: ASC(2)</th>
<th>$Q_3^t$: ASC(3)</th>
<th>$Q_4^t$: ASC(4)</th>
<th>$Q_5^t$: ASC(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_1^t$</td>
<td>0.3333 (0.0118)</td>
<td>0.4082 (0.0193)</td>
<td>0.5896 (0.0120)</td>
<td>0.7427 (0.0189)</td>
<td>0.8964 (0.0470)</td>
</tr>
<tr>
<td>$\phi_2^t$</td>
<td>0.1279 (0.0142)</td>
<td>0.1229 (0.0186)</td>
<td>0.1191 (0.0131)</td>
<td>0.0755 (0.0175)</td>
<td>-0.0285 (0.0690)</td>
</tr>
<tr>
<td>$\phi_3^t$</td>
<td>0.1222 (0.0140)</td>
<td>0.0907 (0.0184)</td>
<td>0.0436 (0.0136)</td>
<td>0.0257 (0.0174)</td>
<td>0.0719 (0.0794)</td>
</tr>
<tr>
<td>$\phi_4^t$</td>
<td>0.1007 (0.0138)</td>
<td>0.0918 (0.0114)</td>
<td>0.0729 (0.0109)</td>
<td>0.0421 (0.0155)</td>
<td>0.0153 (0.0461)</td>
</tr>
<tr>
<td>$\phi_5^t$</td>
<td>0.1513 (0.0114)</td>
<td>0.1693 (0.0119)</td>
<td>0.1298 (0.0097)</td>
<td>0.1402 (0.0158)</td>
<td>0.1291 (0.0469)</td>
</tr>
</tbody>
</table>

Adj-R$^2 = 0.9497$, Prob>F = 0.0000.

* Format in bold means statistically significant at the 1% level.
TABLE IV

The speed of learning from the information content of trades

This table summarizes the estimation of equation (7) by Ordinary Least Squares (OLS) robust to heteroskedasticity and autocorrelation (Newey and West, 1987). The variable \( \tau_t \) is an estimation of the time (in number of events) that quotes need to capture all the information provided by a given trade. This \( \tau_t \) comes from the simulation of the VAR model (3) and is trade-specific. \( D(\tau_t) \) is the series formed with all \( \tau_t \) expressed in real time (seconds). \( V_t \) = 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 \( \Delta q_t \) estimated with a GARCH (1,1) model), and \( D_t, j=[1,...,8] \), are dummy variables that control for deterministic intra-daily patterns.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient*</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_t ) (const.)</td>
<td>721.41</td>
<td>7.987</td>
</tr>
<tr>
<td>( V_t ) (( \delta_t ))</td>
<td>-0.0051</td>
<td>.00018</td>
</tr>
<tr>
<td>( T_t ) (( \delta_t ))</td>
<td>2.1679</td>
<td>.05745</td>
</tr>
<tr>
<td>( R_t ) (( \delta_t ))</td>
<td>-26495.11</td>
<td>3001.8</td>
</tr>
<tr>
<td>( S_t ) (( \delta_t ))</td>
<td>-424.72</td>
<td>22.827</td>
</tr>
<tr>
<td>( QD_t ) (( \delta_t ))</td>
<td>0.4391</td>
<td>.02076</td>
</tr>
<tr>
<td>( D_t ) [9:30 10:00)</td>
<td>-263.98</td>
<td>6.5971</td>
</tr>
<tr>
<td>( D_t ) [10:00 11:00)</td>
<td>-232.71</td>
<td>6.4680</td>
</tr>
<tr>
<td>( D_t ) [11:00 12:00)</td>
<td>-108.26</td>
<td>7.2188</td>
</tr>
<tr>
<td>( D_t ) [13:00 14:00)</td>
<td>7.6262</td>
<td>8.6188</td>
</tr>
<tr>
<td>( D_t ) [14:00 15:00)</td>
<td>-123.29</td>
<td>7.3578</td>
</tr>
<tr>
<td>( D_t ) [15:00 15:30)</td>
<td>-225.56</td>
<td>7.0505</td>
</tr>
<tr>
<td>( D_t ) [15:30 16:00)</td>
<td>-276.85</td>
<td>6.7039</td>
</tr>
</tbody>
</table>

Adj.-R\(^2\): 0.2346  
Prob > F 0.0000

* Format in bold means statistically significant at the 1% level.
FIGURE 1
Intra-daily distribution of the information-asymmetry risk
This 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.

FIGURE 2
Bid-ask spread dynamics after a trade conditional on adverse selection costs
This figure shows the average bid-ask spread weighted by time (not standardized) for each information-asymmetry risk level (ASC(j), j=1,...,5) during the 15 minutes after a trade.
APPENDIX A
The impulse-response function

To perform the simulation of the VAR model (3) we need to define a generating process for $MC_t = (V_t, T_t, S_t, QD_t, R_t)$. It is assumed that each component of $MC_t$ follows a general probabilistic process, exogenous to the model (3). This process is approximated by a linear autoregressive model $AR(p_k)$ like (A.1), where $p_k$, $k={1,...,5}$, must be determined empirically. Model (A.1) is estimated by GLS. Dummy variables are included to control for the deterministic intra-daily components.

$$\varTheta^k_p(L)MC^k_t = \sum_{h=1}^{8} \varphi^k_h D^h_i + u^k_{t,i}. \quad (A.1)$$

Once (2)-(3) and (A.1) have been estimated, the simulation procedure for each of the IBM trades proceeds as follows:

Step #1: Use (A.1), $k={1,...,5}$, to predict the future values of $MC_t$ needed to proceed with the simulation of (3). Assume that $u^k_{t,i} \sim N(\mu^k, \sigma^k_2)$, where $\mu^k$ and $\sigma^k_2$ are estimated through the mean and variance of the GLS residuals of (A.1). The initial conditions $MC^k_{i-1}$ for $i=1,...,p_k$ and $\forall k$, correspond to the values associated to the $p_k$ trades that precede the one simulated.

Step #2: Obtain the impulse-response function (IRF) of (3), $n$ periods into the future, using the predicted values of $MC^k_t$ for $k={1,...,5}$ in step #1. In order to do that, assume that every trade generates a unitary shock ($v_{2_j} = 1$) and is executed after a steady state period, defined by $x_{i-1} = \cdots = x_{i-5} = 0$ and $\Delta q_{i-1} = \cdots = \Delta q_{i-5} = 0$. Obviously, trades are not generally executed after such a steady state period. However, this hypothesis allows isolating the impact associated to a trade from the impact of previous trades. This simulation exercise leads to a realization of the IRF for a given $MC_t$ path. The accumulated impact $T^j_i(\Delta q_i | v_{2_j}, MC_t, D_t)$ for the case of infinite order polynomials appears in equation (A.2). The asterisk means a simulated value, not observed.
\[
\tilde{T}_i(j; \Delta q_t, v_{2,t}, MC_t, D_t) = \sum_{j=1}^{n} a_j \left[ \sum_{i=1}^{n-j} \Delta q^*_{t+i} \right] + \sum_{j=0}^{n} \alpha_j \left[ \sum_{i=1}^{n-j} x^*_{t+i} + v_{2,t} \right] + \\
+ \sum_{j=0}^{n} (B^j_y) \left[ \sum_{i=1}^{n-j} (MC^*_{t+i} + \lambda^q_{t+i} D^*_{t+i}) x^*_{t+i} + (MC_t + \lambda^q_t D_t) v_{2,t} \right].
\]  

(A.2)

Step #3: Repeat steps #1 and #2 10,000 times. The 10,000 estimated conditional values for each step \( n_j (j=1,\ldots,n) \) of the IRF are averaged to obtain the final IRF of the trade,

\[
\bar{T}_i (\Delta q_t \mid v_{2,t}, MC_t, D_t) = \frac{1}{10,000} \sum_{i=1}^{10,000} \tilde{T}_i (\Delta q_t \mid v_{2,t}, MC_t, D_t).
\]  

(A.3)

In this paper \( n=50 \), a period we will evidence to be more than enough for prices to reflect all the information impounded by a trade. The order of the autoregressive process in (A.1) depends on the variable considered, but it is never greater than \( p_k = 5 \) (likelihood ratio tests show that longer lags are not statistically significant for any \( MC^k_t \)). Moreover, as in Hasbrouck (1991a,b), de Jong et al. (1995), and Dufour and Engle (2000), the polynomials in the VAR model (3) are truncated at lag five.
APPENDIX B

Adverse selection costs and market behavior

The following table reports the estimation of the model in equation (8) for several indicators of trading activity, volatility and liquidity. Bold format means that the coefficients are not statistically different from zero. Asterisked values indicate statistically significant differences at the 1% level between the coefficients associated to the same trading hour, in the sense that $c_{k+1} > c_{k}$ for $k=2,…,5$. The coefficient values in the tables are those obtained with the standardized $\zeta^n$ series.

### B.I: Quoted spread

<table>
<thead>
<tr>
<th>Time</th>
<th>ASC(1)</th>
<th>ASC(2)</th>
<th>ASC(3)</th>
<th>ASC(4)</th>
<th>ASC(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 10:00</td>
<td>-0.1826</td>
<td>-0.5012</td>
<td>-0.3221</td>
<td>-0.1427</td>
<td>-0.3279</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>-0.3035</td>
<td>-0.3584</td>
<td>-0.3787</td>
<td>-0.3279</td>
<td>-0.3250</td>
</tr>
</tbody>
</table>

### B.II: Standard deviation of the quote midpoint

<table>
<thead>
<tr>
<th>Time</th>
<th>ASC(1)</th>
<th>ASC(2)</th>
<th>ASC(3)</th>
<th>ASC(4)</th>
<th>ASC(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 10:00</td>
<td>0.3018</td>
<td>0.1945</td>
<td>0.2168</td>
<td>0.1945</td>
<td>0.1945</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>0.3035</td>
<td>0.3035</td>
<td>0.3035</td>
<td>0.3035</td>
<td>0.3035</td>
</tr>
</tbody>
</table>

### B.III: Number of trades

<table>
<thead>
<tr>
<th>Time</th>
<th>ASC(1)</th>
<th>ASC(2)</th>
<th>ASC(3)</th>
<th>ASC(4)</th>
<th>ASC(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 10:00</td>
<td>0.1136</td>
<td>0.5639</td>
<td>0.5947</td>
<td>0.5583</td>
<td>0.5627</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>0.3035</td>
<td>0.3035</td>
<td>0.3035</td>
<td>0.3035</td>
<td>0.3035</td>
</tr>
</tbody>
</table>

### B.IV: Accumulated volume

<table>
<thead>
<tr>
<th>Time</th>
<th>ASC(1)</th>
<th>ASC(2)</th>
<th>ASC(3)</th>
<th>ASC(4)</th>
<th>ASC(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 10:00</td>
<td>0.1826</td>
<td>0.6894</td>
<td>0.0058</td>
<td>0.0111</td>
<td>0.0439</td>
</tr>
<tr>
<td>15:30 – 16:00</td>
<td>0.3851</td>
<td>0.3851</td>
<td>0.3851</td>
<td>0.3851</td>
<td>0.3851</td>
</tr>
</tbody>
</table>

29