Does the Open Limit Order Book Matter in Explaining Long Run Volatility?[*]

Roberto Pascual
Department of Business
Universidad de las Islas Baleares
rpascual@uib.es

David Veredas
ECARES, Université Libre de Bruxelles
CORE, Université catholique de Louvain
dveredas@ulb.ac.be

This version:
June 2006
(First Draft: March 2006)

(Preliminary, suggestions are welcome)

Does the Open Limit Order Book Matter in Explaining Long Run Volatility?

ABSTRACT

This paper evaluates the informational content of an open limit order book by studying its role in explaining long run volatility. We separate liquidity-driven (transitory) volatility from information-driven (long run) volatility using a dynamic state-space co-integration model for ask and bid quotes. We find that changes in immediacy costs, for trades of different sizes, precede posterior fluctuations in long run volatility, even after controlling for the incoming order flow; the book is less informative for large-caps; its informativeness decreases with time aggregation, and the book beyond the best quotes adds explanatory power to the best quotes.

Key words: Limit order book, volatility, electronic order-driven markets, state-space models, price formation, market microstructure.

JEL: G1

1. Introduction

In recent years, several empirical and theoretical studies have examined the informational value of open limit order books (LOBs) in explaining posterior stock returns and the composition of the incoming order flow.[1] With the notable exceptions of Ahn, Bae, and Chan (2001), and Coppejans, Domowitz, and Madhavan (2004), the relationship between price volatility and the state of the LOB has deserved a very limited attention. Ahn *et al* (2001) evidence that volatility declines after an increase in book depth. Similarly, Coppejans *et al* (2004) find that positive liquidity shocks, as measured by the unexpected component of book depth, substantially reduce volatility. These earlier studies do not account for the nature of the observed volatility.

In this paper, we distinguish between the transitory and the informational components in volatility. Transitory volatility is triggered by noise or liquidity trading. It may also be explained by a paucity of limit orders on the book, due to temporary liquidity withdrawals either at the bid or at the ask side. On the contrary, long run volatility relates to the underlying efficient price. It is linked to the trading of informed investors, to the disclosure of public information, and to permanent shocks in the fundamentals of the stock price.[2] Bae et al. (2003) illustrate the importance of distinguishing between these two sources of volatility. They analyze the relationship between volatility and limit order submissions.[3] They decompose the transaction price into an implicit efficient price and a transitory noise, using the time series to estimate

---

[1] See Biais, Hillion, and Spatt (1995); Handa and Schwartz (1996); Parlour (1998); Foucault (1999); Corwin and Lipson (2000); Irvine, Benston and Kandel (2000); Franke and Hess (2000); Coppejans and Domowitz (2002); Cao, Hansch, and Wang (2003); Handa, Schwartz, and Tiwari (2003); Foucault, Kadan, and Kandel (2005); Ranaldo (2004), and Pascual and Veredas (2004), among others.
[2] Ahn et al (2001) implicitly assume their measure of volatility is liquidity-driven. However, there is no justification for such an assumption, since they do not isolate the fraction of volatility due to information flows.
[3] Theoretically, Handa and Schwartz (1996) predict higher liquidity-motivated volatility to increase limit order submissions. Foucault (1999) predicts a similar relationship for information-motivated volatility.

the long run and short run volatility components. They find short run volatility to cause limit order submissions, but no effect is reported for long run volatility.

Our focus is on the role the LOB plays in explaining long run volatility. If the state of the LOB contributes to the price discovery process, then we should observe the volatility of the efficient price to vary with the state of the book. How can the LOB cause informational volatility? It is usually assumed that informed traders always choose to submit market orders, as they are impatient traders endowed with a perishable informational advantage. Limit order traders, however, may also be information-motivated, as theoretically suggested by Kaniel and Liu (2001), Seppi (1997), or Rindi (2003). These papers qualify Copeland and Galai (1983), since periods with high informational volatility might not necessarily be tied to a decrease in limit-order submissions. Recently, Bloomfield, O'Hara, and Saar (2005), and Anand, Chakravarty, and Martell (2005) provide experimental and empirical evidence, respectively, supporting the use of limit orders by informed traders. If limit orders may be information-motivated, then the shape of the LOB may be informative about future information flows and, hence, about long run volatility. In a recent theoretical effort, Foucault, Moinas, and Theissen (2006) develop a model in which limit order traders possess volatility information. In particular, in their model informed dealers have private information on the occurrence of future information events, such as corporate announcements. These dealers adjust their order submission strategies to the level of risk perceived. Uninformed traders learn volatility information from the LOB, and this information is more precise when the traders' identity is not concealed. They conclude that the LOB is a channel for volatility information.

We use high-frequency data on a set of 33 stocks listed in the Spanish Stock Exchange (SSE) in 2000. For each stock, we estimate a structural co-integration model

for ask and bid quotes. We decompose the best quotes into three unobserved components: (a) the common long run component (efficient price); (b) the transitory component of the ask quote, and (c) the transitory component of the bid quote. We allow the short run components of ask and bid quotes to be cross and auto-correlated. In addition, we control for intraday regular patterns in immediacy costs. The efficient price is modeled as a random walk process. As in Hasbrouck (1999), the long run variance (the variance of the efficient price) follows an EGARCH process (Nelson, 1991), which accommodates intraday regularities.

We employ the estimated time series of the long run volatility to evaluate the informational content of different pieces of limit order book information. Using structural vector autoregression (SVAR) models for the long run volatility and the LOB, we show that shocks to long run volatility contemporaneously cause liquidity deterioration by increasing immediacy costs and decreasing quoted depth. More importantly, we report that the LOB does contain information about posterior long run volatility, even when we control for succeeding order flow. Shocks to immediacy costs are followed by an increase in the intensity of information arrival in the short run. Therefore, as predicted by Foucault et al. (2006), variations in immediacy costs measures precede fluctuations in long run volatility. As previously reported by Cao, Hansch, and Wang (2003) and Pascual and Veredas (2004), we confirm there is significant information beyond the best quotes, since immediacy costs measures for hypothetical large-sized trades provide more information than the quoted bid-ask spread. Finally, our findings suggest that the informativeness of the book decreases with time aggregation and market capitalization.

The rest of the paper is organized as follows. In the next section, we present the SSE and the database. In section 3, we present the structural model for ask and bid quotes

and describe the estimation technique. In section 4, we summarize the estimation results of the structural model. In section 5, we evaluate the informational content of the LOB. In section 6, we conclude.

2. Data

Our database consists on 6 months, from July to December 2000, of high frequency data on a set of SSE-listed stocks.[4] All the stocks in the sample are negotiated in the electronic order-driven platform of the SSE, called SIBE. This electronic venue handles the trading activity of the most frequently traded and liquid SSE-listed stocks. Liquidity is provided by an open LOB (there are no market makers). Traders can observe the five best levels of the book, both at the offer and demand sides. The screens are updated in real time. Three basic types of orders are allowed: market, limit, and market-to-limit orders. Market orders walk up or down the book till they are fully executed. Market-to-limit orders are restricted to the best price on the opposite side of the market. Limit orders which are not executed instantaneously get stored on the LOB until they find a counterparty, they are cancelled, or they expire (by default, unexecuted orders expire at the end of the day). A strict price-time priority rule determines the order of execution and storage. Cancellations and modifications are allowed anytime. A trade occurs when an incoming order matches at least one order stored on the opposite side of the LOB. The trading session is continuous from 9:00 a.m. to 5:30 p.m. This period is preceded by a 30-minute opening call auction and followed by a 5-minute closing call auction.

The database includes book and transaction data. We have the five best ask and bid levels of the LOB updated each time the book changes, and time stamped to the nearest hundredth of a second. The book files include the displayed depth and the number of

---

[4] The 2004 "Annual Report and Statistics" of the World Federation of Exchanges classifies the SSE as the 9th world largest stock exchange in terms of capitalization (the 4th among European markets), and the 7th in terms of total value of share trading (the 4th in Europe).

limit orders accumulated at each level.[5] Transaction files update each time the first level of the limit order book changes. It provides information about the accumulated traded volume (in shares) at each update. We apply an algorithm originally developed by Pardo and Pascual (2004) to differentiate between trades, limit order submissions, cancellations, modifications, etc. The algorithm is based on the matching of book and transaction files.

The official market index of the SSE is the IBEX-35, which includes the 35 most liquid and active stocks of the SIBE, weighted by market capitalization. Its composition is regularly revised every semester. Our initial sample is formed by the 35 index constituents during the second semester of 2000. We discard two stocks because they experienced splits. Table I provides some daily descriptive statistics of the remaining 33 stocks. We provide cross-sectional statistics of liquidity and activity for the complete sample, but also separated statistics for the 5 largest and the 15 smallest stocks in terms of market capitalization (as measured by the index weight).

[Table I]

Table I shows that we have a quite heterogeneous sample. The 5 largest (15 smallest) stocks account for 70.76% (7.53%) of the market capitalization, 72.94% (8.9%) of the volume traded, 48.91% (18.02%) of the transactions, and 43.55% (21.48%) of the order flow (market, limit, market-to-limit, and cancellations), of the whole sample. Compared with the 15 smallest stocks, the median daily volume (in shares) for the 5 largest stocks is 35.7 times larger; the euro value of the book depth is 8 times larger; the relative spread is 3.45 times narrower, and they collect 6 times more orders (market, limit, market-to-limit, and cancellations).

---

[5] The SIBE allows submitting partially hidden limit orders (see Pardo and Pascual, 2004).

We use the abovementioned book files to compute summary measures that will characterize the state of the LOB at each point in time. In particular, in the following sections we will consider the following variables:

(a) The bid-ask spread (S_t).

(b) The displayed depth at the best ask (D_t^a) and bid (D_t^b) quotes.

(c) The quoted depth beyond the best quotes, defined as the accumulated depth at up to k ticks from the quote midpoint. We distinguish between depth at the ask side ($BD_t^a(k)$) and at the bid side ($BD_t^b(k)$). This measure is the minimum volume needed to move the price by k ticks, a natural measure of liquidity.

(d) Order imbalances in the limit order book are computed as $AsD_t = \left| D_t^a - D_t^b \right|$ for the best quotes, and $AsBD_t(k) = \left| BD_t^a(k) - BD_t^b(k) \right|$ for the book beyond the best quotes.

(e) We follow Irvine *et al* (2000) in defining our bi-dimensional liquidity measure as,

$$BLM_t(s) = \frac{PI_t^a(s) - PI_t^b(s)}{Q_t},$$
[1]

where $PI_t^a(s)$ ($PI_t^b(s)$) is the price impact of a fictitious buyer-initiated (seller-initiated) trade with size s times the normal market size (NMS). The NMS is computed as the median trade size per stock and month. Finally, Q_t represents the quote midpoint. The measure in [1] could be understood as the immediacy costs of a round-trip for a medium-sized order.

3. The dynamic state-space model for ask and bid quotes

In this section, we introduce a microstructure structural model for the bid and ask quotes that features short run and long run volatility effects. The model shares some similarities with the dynamic model proposed by Hasbrouck (1999), but there are remarkable disparities. Most notably, we do not explicitly model price discreteness, as Hasbrouck did, so as to preserve the linearity of the model.[6]

The model builds on the traditional microstructure price decomposition that splits the quoted prices into a "true" long run efficient price and a transitory non-informative component (e.g., Hasbrouck, 1996). In our particular case, the ask and bid quotes are given by,

$$\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \beta \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} S_{at} \\ S_{bt} \\ m_t \end{pmatrix}, \qquad [2]$$

where a_t and b_t stand for the ask quote and the bid quote, respectively. Quoted prices are co-integrated, with theoretical co-integration vector $(1,-1)$ (e.g., Engle and Patton, 2004). The β coefficient captures the average difference between the two quoted prices, which must be always positive. Hence, we expect β to be positive and close to half the average bid-ask spread. The observed quoted prices are a function of three unobservable factors, S_{at} and S_{bt}, which are the transitory components of ask and bid quotes that move the quoted prices away from the efficient price (m_t). These transitory components are due to microstructure frictions, including the tick size.

---

[6] Hasbrouck (1999) models discreteness using an asymmetric rounding of the ask and bid quotes around the efficient price. Although quotes are discrete in nature, we believe that the progressive conversion into decimals in markets around the world has significantly reduced the importance of the practical implications of ignoring discreteness. In the particular case of the SSE, the minimum tick is 0.01€ for stocks with prices below 50€.

The dynamics of the unobservable components of ask and bid quotes are modeled as,

$$\begin{pmatrix} S_{a,t} \\ S_{b,t} \\ m_t \end{pmatrix} = \begin{pmatrix} \phi_a & \phi_{ab} & 0 \\ \phi_{ba} & \phi_b & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} S_{a,t-1} \\ S_{b,t-1} \\ m_{t-1} \end{pmatrix} + \begin{pmatrix} g_t \\ -g_t \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \end{pmatrix}. \qquad [3]$$

As it is usually assumed in microstructure structural models of price formation, the efficient price m_t follows a random walk process. In contrast, the transitory components are stationary, and follow a bivariate first order autoregressive process. We allow for lagged causality between the transitory components. Thus, a shock in $S_{b,t}$ ($S_{a,t}$) may affect $S_{a,t}$ ($S_{b,t}$) the next period if ϕ_{ab} (ϕ_{ba}) is statistically different from zero. The function g_t is linear in a set of J dummy variables $D_{t,j}$ for the time interval, which are meant to capture the deterministic intraday pattern in the bid-ask spread.[7]

$$g_t = \sum_{j=1}^{J} \phi_j D_{t,j}. \qquad [4]$$

The vector of innovations in [3] is assumed to be jointly normally distributed,

$$\begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \end{pmatrix} \square\, N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_t \right), \qquad [5]$$

with the following time-varying variance-covariance matrix,

$$\Sigma_t = \begin{pmatrix} \sigma_{a,t}^2 & 0 & 0 \\ 0 & \sigma_{b,t}^2 & 0 \\ 0 & 0 & \sigma_{m,t}^2 \end{pmatrix}. \qquad [6]$$

---

[7] In particular, we split the SSE trading session into the following time intervals: [9:00, 9:30), [9:30, 10:00), [10:00, 13:00), [13:00, 15:00), [15:00, 15:30), [15:30, 16:00), [16:00, 16:30), [16:30, 17:00), [17:00, 17:30). With this particular splitting, we attempt to capture the widely evidenced particularities of the initial and final intervals of the trading session in financial markets, but also the effects associated with the opening of the US markets at 15:30 (Spanish time). So as to get a more parsimonious model, we impose the same intraday deterministic pattern for the transitory components of ask and bid quotes. Preliminary estimations using a more parameterized model suggest this restriction is not a very strong one.

Certainly, normality is a less than ideal assumption. In our setting, however, as far as the model is linear, this distributional assumption allows us to estimate the factors (m_t, $S_{a,t}$, and $S_{b,t}$) using the Kalman filter, and the parameters using the error prediction decomposition of the Kalman filter (see Harvey, 1992). Moreover, using Pseudo-Maximum Likelihood (PML) arguments in a Quadratic Exponential Family setting (the Gaussian distribution belongs to this family), the estimates are consistent, albeit not efficient, under distributional misspecification, as far as the conditional means and variances are correctly specified. More specifically, if the parameters of interest are in the conditional mean and the conditional variance, consistency is guaranteed as far as the assumed density belongs to the Quadratic Exponential Family, as it is in our case (see Gouriéroux, Monfort, and Trognon, 1984).

The variances of the short run components $S_{a,t}$ and $S_{b,t}$ are given by $\sigma_{a,t}^2$ and $\sigma_{b,t}^2$. Rather than imposing homoskedasticity, as in Hasbrouck (1999), we assume them to be time-varying in an entirely deterministic way, showing intraday regular patterns,

$$\sigma_{i,t}^2 = \exp\left(\sigma_i^2 + \sum_{j=1}^{J} \phi_j^i D_{t,j} \right), \ i = \{a,b\}. \tag{7}$$

The exponential form in [7] is used for convenience, so as to guarantee the non-negativeness of the conditional variance, since some ϕ_j^i could be negative.

In contrast, long run volatility ($\sigma_{m,t}^2$) is allowed to have deterministic and dynamic components, which are captured, following Hasbrouck (1999), by an EGARCH model,

$$\sigma_{m,t}^2 = \exp\left(\alpha_0 + \alpha_1 \ln\left(\sigma_{m,t-1}^2 \right) + \alpha_2 \xi_{m,t-1} + \alpha_3 \left[\left| \xi_{m,t-1} \right| - \sqrt{2/\pi} \right] + \sum_{j=1}^{J} \phi_j^m D_{t,j} \right), \tag{8}$$

where $\xi_{m,t} = \varepsilon_{m,t} / \sigma_{m,t}$ is the standardized shock. Model [8] includes an intercept (α_0), a persistence parameter (α_1), a sign or leverage parameter (α_2), and a magnitude parameter (α_3). Thus, α_2 captures the asymmetric effects of good and bad news of prior shocks, while α_3 captures the effect of the size of these same shocks. Finally, deterministic intraday patterns in long run volatility are captured by the set of dummy variables previously defined.[8] The exponential form in [8] guarantees the non-negativeness of the conditional variance, as some of the ϕ_j^m may take negative values.

Equations [2] to [8] form a linear state space model. The standard Kalman filter, however, cannot be applied to estimate the model's factors and parameters because of the EGARCH assumption in [8]. Since the efficient price is unobserved, the volatility model in [8] is a function of the squares of past unobserved shocks, and the standard Kalman filter is unfeasible. To circumvent this problem, we rely on Harvey, Ruiz, and Sentana (1992), who propose to augment the state space and treat the error term of the efficient price as another state variable to be estimated jointly with $S_{a,t}$, $S_{b,t}$ and m_t. Once the state space has been augmented, the estimation of the factors with the Kalman filter, and of the parameters with the error prediction decomposition, is possible. We refer to the Appendix for a brief explanation of the augmented state space model.

4. Estimation of the state-space model

We proceed next to summarize the estimation of the state space model in [2]-[8] for the 33 Spanish stocks in our sample. The time series of ask and bid quotes are computed in 5-minute intervals using the last quotes in each interval. Table II provides cross-

---

[8] We allow both transitory and informational volatility to show intraday regular patterns. While deterministic components in long run volatility would reveal concentration in information revelation at particular intervals of the trading session, regular patterns in short run volatility are not information-driven.

sectional statistics on the estimated parameters in equations [2]-[3] and [7]-[8]. Panel A in Table II provides statistics for the whole sample, but also for the 5 largest and 15 smallest stocks by market capitalization.[9] Panel B provides correlations between the estimated parameters and daily statistics about liquidity and activity, as previously described in Table I.

[Table II]

In concordance with the descriptive statistics in Table I, Panel A in Table II shows that the estimated average bid-ask spread (2β in [2]) for the 5 largest stocks in narrower than for the 15 smallest stocks. The median estimated spread is around 4 ticks (0.04134€) for the whole sample, 1.6 ticks for the largest stocks, and 4.6 ticks for the smallest stocks. In addition, Panel B in Table II evidences that $2\hat{\beta}$ decreases when the stock is liquid, as measured by the relative bid-ask spread and the book depth. Moreover, the estimated β is smaller among frequently traded stocks, with intense order flow, and reduced volatility. Indeed, $2\hat{\beta}$ approximates quite precisely the true median bid-ask spread, as it is shown in Figure 1; the cross-sectional correlation between both measures is 0.99.

[Figure 1]

Panel A in Table II evidences a high degree of persistency in the short-term components of prices, $S_{a,t}$ and $S_{b,t}$. For the whole sample, the cross-sectional median of the estimated ϕ_a (ϕ_b) in [3] is 0.789 (0.6269). These parameters decrease with market capitalization, with the median ϕ_a (ϕ_b) being equal to 0.6069 (0.4756) for the 5 largest stocks and 0.8174 (0.6773) for the 15 smallest stocks. Panel B in Table II shows they

---

[9] The individual-security estimated coefficients of the state-space model, which are summarized in Table II, generally are statistically significant at the 1% level. Exceptionally, for some stocks we obtain non-significant coefficients for a few of the intraday dummies used in equations [7] and [8] to capture the intraday regular patterns in volatility, implying that these periods do not differ from the benchmark time.

are also strongly negatively correlated with liquidity. The most active stocks, either in terms of volume, trades, or order flow, also experience lower persistency. These implies that the largest, the most liquid, and the most active stocks, also exhibit the desirable property of being the most resilient.[10] Kyle (1985) considers resiliency as a third dimension of liquidity, together with immediacy costs and depth. Therefore, it is reasonable to expect least liquid and active stocks to be the ones for which unusual deviations from the efficient price due to uninformative shocks take longer to be reverted. Moreover, Biais, Hillion, and Spatt (1995) show that, in order driven markets, limit order traders provide liquidity when it is valuable for the marketplace, a finding corroborated by Ahn *et al* (2001) and Ranaldo (2004), among others. Since the costs of submitting limit orders increases with the expected time to execution (e.g., Lo, MacKinlay, and Zhang, 2002), uninformative shocks should be reversed earlier when the stock is highly active and, therefore, the average time to execution is shorter. We also find $S_{a,t}$ and $S_{b,t}$ to cause each other in the short run, with the estimated ϕ_{ab} (ϕ_{ba}) in [3] being positive. This implies that transitory components of ask and bid quotes move together, though not necessarily in a symmetric way.[11]

Figure 2 represents the estimated intraday regular patterns in the transitory components of ask and bid quotes, given by the estimated parameters ϕ_j of the g_t deterministic function in [4]. The values we report are the cross-sectional median of $2\hat{\phi}_j$ for each time interval. We compare the estimated deterministic pattern with the realized cross-sectional median level of the bid-ask spread during the same time intervals of the trading session. Figure 2 shows the estimated regular patterns fit reasonably well the observed intraday deterministic behavior of the bid-ask spread. The spread achieves its

---

[10] Resiliency refers to the speed with which prices recover from a random, uninformative shock.
[11] Escribano and Pascual (2006) report asymmetric dynamics in ask and bid quotes for NYSE-listed and SSE-listed stocks.

maximum level at the beginning of the session and sharply decays in one and a half hours. Minimum levels of the spread are observed toward the end of the day, with a slight increase right after the opening of the US markets at 15:30 Spanish time.

[Figure 2]

Panel A in Table II provides the cross-sectional median level of the transitory volatility (σ_a^2 and σ_b^2 in the RHS of [7]) during the control interval of the trading session (from 1:00 p.m. to 3:00 p.m.). Results are similar for ask and bid quotes. Transitory volatility is larger for the 15 smallest stocks than for the 5 largest stocks. This finding should not be surprising since transitory volatility is associated with the paucity of liquidity. Indeed, Panel B in Table II shows that σ_a^2 and σ_b^2 both decrease with the average liquidity and activity of the stock.

As for the long run volatility in [8], the persistence parameter (α_1) is larger among the 5 largest ($\alpha_1 = 0.9759$) than among the 15 smallest ($\alpha_1 = 0.8972$) stocks. Similarly, Panel B in Table II reports a positive cross-correlation between α_1 and both liquidity and activity proxies. Since information comes in clusters, informational volatility also progresses in clusters of high and low levels. Our findings point toward information flow intensity being more stable among the largest stocks in the sample, since volatility clusters last longer for these stocks. Consistently, we obtain that the estimated magnitude parameter (α_3) is lesser among the large caps ($\alpha_3 = 0.1578$) than among the small caps ($\alpha_3 = 0.3625$), meaning that the formers tend to be less responsive to an informative shock of any given size. Panel B in Table II corroborates this finding reporting that α_3 is larger among illiquid and infrequently traded stocks. These findings are consistent with the very well known fact that smaller, less liquid, and less active stocks suffer from higher information asymmetry risk (e.g., Stoll, 2000).

Finally, the leverage parameter in [8] (α_2) is smaller in magnitude than the other parameters in the EGARCH model, but still statistically significant. Table II shows that α_2 tends to be negative for large, active, and liquid stocks. As usually documented, negative shocks cause larger fluctuations in the efficient price than positive shocks. For the smaller, less liquid and less active stocks, however, α_2 tends to be positive, implying that good news induce a larger variance than bad news.

Figure 3 represents the cross-sectional median of the intraday regular patterns in transitory and informational volatility, as given by the estimated ϕ_j^i parameters in [3] and the ϕ_j^m parameters in [8], respectively. Figure 3 represents the deviation of the estimated parameters with respect to the control interval (from 13:00 to 15:00).

[Figure 3]

This figure reveals important differences in the estimated regularities of short run and long run volatility. While transitory volatility regularly reaches its maximum level at the initial intervals of the session, informational volatility achieves its minimum during the first halt-hour of the session. Both components of volatility increase towards the end of the day, right before the opening of the US markets. Informational volatility achieves its regular maximum in the last half-hour of the trading session. This clustering of informative shocks is consistent with the high α_1 parameter found in the EGARCH model [8]. Recently, Pascual, Pascual-Fuster, and Climent (2006) have shown that the contribution of the NYSE to the price discovery process of the Spanish cross-listed stocks during the two-hour overlapping interval between both markets is negligible. Nonetheless, the opening of the US market is regularly followed by a sharp increase in the SSE trading activity and order flow (see Pardo and Pascual, 2004). The long run volatility patterns in Figure 3 are consistent with this final highly active two-hour period

of the SSE session being also the most contributive to price discovery. Besides, the initial peak in short run volatility coincides with the least liquid trading period, the opening of the session.

Panel C in Table II provides cross-sectional correlations between liquidity, and activity measures and the average of the estimated transitory components of ask and bid quotes, $S_{a,t}$ and $S_{b,t}$ in [2], and the estimated long run and short run volatilities, $\sigma_{m,t}^2$ in [8], and $\sigma_{a,t}^2$ and $\sigma_{b,t}^2$ in [7], respectively. Panel C reinforces the results in previous panels in Table II. Short run components of ask and bid quotes are larger and more volatile in median terms among the smallest, less liquid, and less active stocks in our sample. That market frictions are negatively related to measures of trading activity, such as volume or number of trades, are empirical regularities (e.g., Stoll, 2000). More intriguing is the finding that $\sigma_{m,t}^2$ is positively related with trading activity, as measured by the number of trades and the intensity of the order flow. A large part of market microstructure research builds on the notion that trades convey new information about the long run value of the stock.[12] Therefore, trades move prices and cause long run variance. For the NYSE case, Hasbrouck (1991b) founds trading to contribute more in both relative and absolute terms to the long run variance of the small caps. Moreover, he reports that the long run variance decreases with market capitalization. Our findings do not agree with the NYSE case: for the 5 largest stocks in our sample, the median of $1000\sigma_{m,t}^2$ is 0.63, while for the 15 smallest stocks is 0.11; for the 5 most frequently traded stocks, in terms of number of trades, the median $1000\sigma_{m,t}^2$ is 0.75, while for the 15 less frequently traded stocks is 0.33. In order to make a careful study on the motives of this apparent inconsistency, we would require a larger and matched sample of NYSE

---

[12] Models for the estimation of adverse selection costs using structural models of price formation are based on this basic principle (e.g., Huang and Stoll, 1997).

and SSE stocks, and a deeper comparative study of the alternative methodologies used in both papers, something that is out of the scope of this paper.

5. The informativeness of the limit order book

In this section, we use the time series of the informational volatility estimated through the state space model in previous sections to evaluate the informational content of the LOB.[13] In particular, we use multivariate time series models to check whether different pieces of LOB information are able to explain posterior variations in informational volatility. Admittedly, it could be argued that the variables of the LOB should have been introduced in equation [8] as part of the EGARCH specification. However, doing so raises two problems. The first one, rather conceptual, is that we are not interested in how the LOB information influences on the estimation of the informational variance, but in how the LOB affects the unobserved informational variance. A one-step procedure would require estimating the long run volatility conditional on its own past and the state of the LOB, which would bias our empirical analysis. The second inconvenience is that we would be facing a potential problem of endogeneity by introducing contemporaneous LOB variables to explain long run volatility. Instead, we use structural vector autoregressive models (SVAR hereafter) to study the dynamics between the book and the informational volatility.

The variables we consider to describe the state of the LOB are the ones defined in section 2: the bid-ask spread (S_t); the displayed depth at the best ask (D_t^a) and bid (D_t^b) quotes; the accumulated depth at up to four ticks from the quote midpoint ($BD_t^a(4)$, $BD_t^b(4)$); the order imbalance on the first level of the LOB (AsD_t); the order

---

[13] The fitting of the state space model is generally worse for the first observations in the time series due to the initialization of the filter. We therefore discard the fist 15 days.

imbalance beyond the best quotes ($AsBD_t(4)$), and the bi-dimensional liquidity measure in [1] for a trade of size equal to 2 times the NMS ($BLM_t(2)$).[14] We are only interested in the informative component of these variables. Therefore, we remove their deterministic, and therefore predictable, component by regressing them against the set of dummy variables in equations [7].[15] Moreover, we would like each of these pieces of LOB information to be additive, in the sense of not providing redundant information. So as to identify the dynamics of each variable that are not responding to the other LOB variables, we estimate the following VAR model,[16]

$$y_t = \gamma_0 + \sum_{j=1}^{3} \gamma_j' y_{t-j} + w_t,$$ [9]

where y_t is the vector of non-deterministic components of the LOB variables. We stack together the time series of the 33 stocks in the sample. We do not allow any lag to reach back to the previous day, implying that we drop the first three observations of each day. Table III summarizes the estimation of model [9] by least squares for time series in a 5-minute periodicity. Panel A provides the estimated coefficients. For each equation and variable, we sum the coefficients at all lags whenever they are statistically significant (at least) at the 5% level. Panel B reports the residual correlation matrix.

[Table III]

We find a negative relationship between immediacy cost and quoted depth. An increase in either D_t^i or $BD_t^i(4)$ causes posterior decreases in S_t and $BLM_t(2)$, and vice versa. The negative relationship between both liquidity dimensions was originally

---

[14] Our main conclusions do not vary if we consider larger values for k and s to compute $BD_t^i(k)$ and $BLM_t(s)$.

[15] Results are not reported because of space limitations, but they are available upon request from the authors.

[16] We truncate the VAR model at lag 3. Additional lags are not statistically significant for almost all the dependent variables.

reported by Lee, Mucklow, and Ready (1993). Moreover, imbalances between bid and offer depth cause posterior increases in immediacy costs. This finding is at odds with Handa, Schwartz, and Tiwari (2004), which predict the bid-ask spread in order-driven markets to achieve its maximum when the imbalance between buyers and sellers is zero. Panel B in Table III shows that some of the w_t residuals are strongly contemporaneously correlated, most notably $BLM_t(2)$ and S_t.

From now on, we will use the residuals w_t in [9] as LOB variables. Since they have been filtered by deterministic intraday components and expected components given the past values of the LOB, these residuals could be interpreted as the unpredictable component of the LOB. If the LOB were informative, information should reside in w_t.

The dynamic response of the information-driven volatility to an innovation in the LOB is measured by the coefficients of the following VAR model,

$$z_t = \begin{pmatrix} \tilde{\sigma}_{m,t} \\ w_t \end{pmatrix} = B_0 + \sum_{j=1}^{p} B_j ' z_{t-j} + C' x_t + \varepsilon_t, \qquad [10]$$

with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma$, $E(\varepsilon_t \varepsilon_s') = 0 \; \forall \; s \neq t$, and x_t being a vector of exogenous variables we will make explicit latter. The first component of z_t, $\tilde{\sigma}_{m,t}$, is the non-deterministic component of the estimated long run standard deviation. From [8],

$$\tilde{\sigma}_{m,t} = \sqrt{\frac{\hat{\sigma}_{m,t}^2}{\exp\left(\sum_{j=1}^{J} \hat{\phi}_j^m D_{t,j}\right)}} \; . \qquad [11]$$

A problem with VAR analysis is that the variance-covariance matrix of the innovations (Σ) is not restricted to be a diagonal matrix. Therefore, a shock to one variable provides information about the innovations to other variables. This implies that

causal interpretations of, for example, simple impulse response functions are not possible. Nonetheless, the VAR in [10] could be considered as the reduced form of a more general structural VAR (SVAR). Let P be a matrix such that $\Sigma = PP'$. Then, $e_t = P^{-1}\varepsilon_t$ are orthogonal unit-variance disturbances ($E(e_t e_t') = I_k$). Pre-multiply both sides of the VAR model [10] by a squared matrix A of contemporaneous correlations, with $AP = V$. It follows that we can represent the VAR model in [10] as,

$$Az_t = A_0 + \sum_{j=1}^{p} A_j z_{t-1} + F'x_t + v_t, \qquad [12]$$

where $A_j = AB_j$, $F = AC'$, and $v_t = Ve_t = A\varepsilon_t$, with $E(v_t v_t') = VV'$.

By imposing the convenient restrictions in matrix A and V, we can isolate the shocks to the LOB from the shocks to the long run volatility. In our particular application, we follow Coppejans, Domowitz, and Madhavan (2004) in setting the identification restrictions. These authors do not allow liquidity to contemporaneously affect returns. They argue that liquidity is a function of bid and offers, which naturally precede transactions, and therefore returns. Ranaldo (2004) evidenced that the state of the LOB determines posterior order flow composition. Therefore, the state of the LOB naturally precedes cancellations, and market and limit order submissions. Since the market infers about the true value of the stock from the order flow composition, the LOB should affect long run volatility in the next period. In contrast, our specification permits long run volatility to contemporaneously cause the LOB, since long run volatility is associated with the arrival of new information, and liquidity is known to be inversely related to adverse selection costs. Moreover, informational volatility influences the choice between market and limit orders (Foucault, 1999), and therefore the shape of the

LOB. According to these arguments, the matrix A for an hypothetical model with 4

variables, $\tilde{\sigma}_{m,t}$ and three LOB measures, would be of the form,

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ a_{31} & 0 & 1 & 0 \\ a_{41} & 0 & 0 & 1 \end{bmatrix}, \qquad [13]$$

with a_{ij} being unrestricted parameters.

Additionally, a shock to either LOB variable is likely derived from the same source

of information than contemporaneous shocks to other LOB variables. Therefore, our

identification restrictions allow the shocks in different LOB variables to be

contemporaneously correlated. Thus, matrix V for the hypothetical 4-variable model

would be of the form,

$$V = \begin{bmatrix} v_{11} & 0 & 0 & 0 \\ 0 & v_{22} & 0 & 0 \\ 0 & v_{32} & v_{33} & 0 \\ 0 & v_{42} & v_{43} & v_{44} \end{bmatrix}, \qquad [14]$$

with v_{ij} being unrestricted parameters.

The reduced (VAR) form [10] is a seemingly unrelated regression model. Since the

model has the same explanatory variables in each equation, estimating the coefficients

equation by equation by least squares generates the maximum likelihood estimates (e.g.,

Greene, 2003). The residuals of the VAR model [10] could then be used to estimate the

variance-covariance matrix of ε_t (Σ). With the restrictions [13]-[14], model [12] is

exactly identified. Therefore, the coefficients of the structural model [12] could be

recovered from the coefficients of the reduced form [10].

As previously discussed, the LOB influences the incoming order flow composition, and the market could learn from the particular mix between market orders and limit orders. However, this would not necessarily mean the book is informative, since the market would be learning from the order flow, but not from the LOB. To account for the order flow between *t-1* and *t,* the vector x_t of control variables in [10] and [12] includes the number of trades completed (T_t) and the total number of orders submitted (O_t) in each time interval. The second variable includes limit orders, market orders, market-to-limit orders, and cancellations. We exclude order modifications. By including these control variables, we intend to evaluate whether the LOB is informative per se. As with the LOB variables, we remove the deterministic components of T_t and O_t using the intraday dummy variables in [7].

In the next subsections, we report our main findings under alternative model specifications. First, we summarize the estimation of models [10] and [12] with time series aggregated in 5-minute intervals. Then we check the robustness of our findings to time aggregation and market capitalization.

5.1. SVAR estimates with five-minute time series

Table IV summarizes the estimation of the SVAR model [12] with a 5-minute periodicity. The time series of all the stocks are pooled together into a single estimation. We rescale each time series dividing by its average value per stock. Then, we stack together the time series of the 33 stocks in the sample. In this manner, we can evaluate the informativeness of the LOB by directly checking the significance of the pooled regression coefficients. We consider three alternative definitions of the vector z_t of dependent variables. The "best quotes" model includes liquidity measures based on the best ask and bid quotes, that is, $z_t' = (\tilde{\sigma}_{m,t}, S_t, D_t^a, D_t^b, AsD_t)$. The "book" model

considers liquidity measures based on the available five best book levels of ask and bid quotes; in this case, $z'_t = (\tilde{\sigma}_{m,t}, BLM_t(2), BD^a_t(4), BD^b_t(4), AsBD(2))$. Finally, the "best + book" model includes all the LOB variables in the former models. All models are truncated at lag one.[17]

[Table IV]

Panel A in Table IV reports the estimated free coefficients of the matrix A in [12] for each of the three models considered. This panel shows that an increase in long run volatility causes a contemporaneous decrease in the liquidity supply by the LOB. A positive shock to long run volatility increases immediacy costs, as measured by both S_t and BLM_t, and decreases the accumulated book depth, BD^a_t and BD^b_t. Panel A also reports that increases in informational volatility contemporaneously enhance the imbalance between the ask side and the bid side of the book. These findings are in harmony with the widely available evidence on the negative relationship between adverse selection costs and liquidity.

Panel B in Table IV reports the estimated coefficients in the $\tilde{\sigma}_{m,t}$ equation of the reduced form model [10].[18] We also provide the Granger causality test for each liquidity measure and also for all the LOB information. The best quotes model shows that wider bid-ask spreads lead to periods of greater informational volatility. More remarkable, this information content in the bid-ask spread persists even when we control for the incoming order flow, implying that investors grasp additional information from the state of the LOB. Changes in quoted depth at the best ask and bid quotes, however, are not

---

[17] Formal lag-selection criteria are not very conclusive regarding the optimal lag length. We have considered specifications up to three lags and our main conclusions are the same. Therefore, we opt for the most parsimonious specification.

[18] Results from the other equations of each model are omitted for space limitations, but they are available upon request from the authors.

found to signal changes in the intensity of information arrival. Indeed, the Granger causality tests for this model clearly indicate that only the quoted bid-ask spread has some explanatory power on posterior long run volatility. The book-based model points to similar conclusions than the best-quotes model. The bi-dimensional liquidity measure in [1] is highly statistically significant, positively causing the intensity of information arrival. Although in this model depth measures are found to be statistically significant (BD_t^a, BD_t^b), causality tests indicate that their contribution in explaining posterior informational volatility is negligible.

Finally, the estimates of the best-plus-book model, in the last column in Panel B, suggest that the market gleans additional information from the secondary steps of the book. In all models, causality tests reject the null that the LOB variables do not cause informational volatility, but this rejection is stronger for models that include book-based proxies of liquidity. Our findings are in line with Cao *et al*'s (2003) findings, who report the book beyond the best quotes to account for a remarkable part of the information share of the complete LOB. Consistently, among all the liquidity measures considered, BLM_t happens to be the most important in explaining long run volatility. Regarding the sign of the relationship between the LOB and the long run volatility, our findings are not conclusive. While the book-based measures indicate that larger immediacy costs, lower depth, and higher imbalances on the book anticipate periods of more intense information arrival, the coefficients of the best-quotes-based variables suggest just the opposite. Nonetheless, as previously remarked, book-based liquidity measures dominate.

The LOB is updated constantly due to the permanent flow of orders. Consequently, its shape at a particular point in time could be informative only about the very short run. If this were the case, time aggregation could bias our analysis towards the conclusion

that the book is less informative than it really is. This line of argument could be particularly appropriate for the case of the largest stocks, which usually are the most frequently traded too. As a matter of robustness, in the next sections we repeat our former study, first, for the 5-largest and 15 smallest stocks in our sample, and, then, with time series computed over 1-minute intervals.

5.2. SVAR estimates conditional on market capitalization

Table V summarizes the estimation of the SVAR model [12] for the 5 largest stocks and the 15 smallest stocks in the sample with a 5-minute resolution. As in Table IV, we provide the estimated free parameters of the matrix of contemporaneous relationships in Panel A, and the estimated coefficients of the reduced form [10] for three different model specifications in Panel B.

[Table V]

Table V reinforces the findings reported in Table IV for the complete sample. Panel A shows greater long run volatility contemporaneously causing increases in illiquidity, both among the largest and among the smallest stocks, and most notably when the book-based liquidity measures are considered. In addition, Panel B in Table V shows that the lagged causality running from liquidity to informational volatility is robust across subsamples. For both the small-caps and the large-caps, an increase in $BLM_t(2)$ precede an increase in the intensity of information arrival. Causality tests in Table V strengthen the idea that $BLM_t(s)$ could be a good summary measure of the informational content of the LOB, especially among the smallest stocks. The incidence of depth-based measures is again imprecise, since the level of significance and the direction of the relationship vary with the model specification.

Granger causality tests reject, for all model specifications and for the two subsamples, the null that the shape of the book does not provide relevant information about posterior informational volatility, but with more strength for the least frequently traded stocks. Franke and Hess (2000) conclude the book is informative only during periods of low information intensity. Likewise, our findings in Table V suggest that, once we control for the incoming order flow, the LOB is more informative among the least frequently traded, probably because the book is steadier.

5.3. SVAR estimates with one-minute time series

Table VI reports the results of estimating the SVAR model [12] with a 1-minute resolution. As in previous tables, we provide the pooled contemporaneous coefficients of the SVAR model in Panel A, and the pooled regression coefficients of the reduced form [10] in Panel B.

[Table VI]

Regarding the contemporaneous coefficients of the SVAR model in [12], Panel A reports similar findings to those shown in former model specifications. Higher informational volatility contemporaneously causes liquidity by increasing immediacy costs and, to a lesser degree, by decreasing quoted depth. Moreover, Panel B in Table VI reports significant lagged causality running from the LOB variables to the long run volatility. In the "best quotes" model and the "book" model specifications, increases in immediacy costs precede periods of higher information volatility, consistently with the theoretical predictions in Foucault et al. (2006). Granger causality test reject the null that the LOB information does not contain volatility information with more strength in the "book" model than in the "best quotes" model, suggesting there is some additional information contained in the secondary levels of the LOB, consistently with Cao et al.

(2003) empirical findings. Finally, the "best + book" model shows that the piece of LOB information with higher explanatory power is the bi-dimensional measure in equation [1].

Table VI confirms our findings are robust to time aggregation. Indeed, the rejection of the null of no causality is more dramatic in Table VI than in Table IV, which suggests the dynamics between the LOB and the long run volatility being better captured with shorter time resolutions. We have also estimated model [12] with a 1-minute resolution for the 5 largest and 15 smallest stocks separately. Our findings do not remarkably differ from those reported in Tables V and VI.[19] Knowing the state of the LOB at a given point in time is more informative about the intensity of information arrival for the least frequently traded stocks.

6. Conclusions

In this paper, we use data on a set of 33 Spanish stocks to evaluate whether open limit order books (LOBs) contain relevant information for price discovery. In particular, we examine the dynamic relationship between the state of the LOB and the long run (informational) volatility of the stock, that is, the volatility of the efficient price.

We estimate informational volatility using a dynamic co-integration state-space model for ask and bid quotes. In this model, ask and bid quotes are decomposed into three unobservable factors: the efficient price, and the transitory component of each quote. The short run components are modeled as a stationary bivariate first order autoregressive process. The efficient price is modeled as a heteroskedastic random walk process, common to both ask and bid quotes. The variance of the innovations in the efficient price is stochastic, and follows an Exponential GARCH (EGARCH) model.

---

[19] We do not provide these findings because of space limitations, but they are available upon request from the authors.

27

We estimate the unobserved components of ask and bid quotes by the Kalman filter, and the model parameters by Pseudo Maximum Likelihood. Finally, we use multivariate time series models to study the dynamic relationship between different pieces of LOB information and the estimated long run volatility.

Our main conclusion is that the state of the LOB explains posterior fluctuations in the intensity of the information flow, as recently suggested in a theoretical paper by Foucault et al. (2006). We find that immediacy costs measures computed for imaginary trades of different sizes are positively related with higher levels of informational volatility in the very short run. The shape of the LOB is informative even when the incoming order flow and trading activity are taken into account. Despite our conclusions are robust to time aggregation, a comparison between models with 1-minute and 5-minute resolution suggests that the dynamics between liquidity and long run volatility are better captured at very short time frequencies. We also show the LOB is informative for both the most frequently traded and the least frequently traded stocks in our sample, even though causality is accepted with more strength for the latter. This finding is consistent with more unsteady LOB books being less informative. Finally, book-based liquidity measures have higher explanatory power than best-quotes-based liquidity measures in explaining informational volatility. This finding is in line with Cao et al's (2003) conclusions: the information derived from the secondary steps of the book adds explanatory power to the information contained in the best quotes.

This paper complements previous empirical studies addressing the same issue but with different methodological approaches (e.g., Ranaldo, 2004; Harris and Panchapagesan, 2005, and Cao et al., 2003). All of them conclude that the LOB is informative. Moreover, our findings are in harmony with recent empirical studies (see Bloomfield et al, 2005, and Anand et al, 2005) showing that limit order traders may be

information-motivated. In particular, as theoretically suggested by Foucault et al. (2006), limit order traders may posses volatility information, that is, they may have advanced information about the occurrence of future informative events, such as corporate announcements.

References

Ahn, Hee-Joon, Kee-Hong Bae, and Kalok Chan, 2001, Limit orders, depth and volatility: evidence from the stock exchange of Hong Kong, *Journal of Finance*, 56, 2, 769-790.

Anand, Amber, Sutago Chakravarty, Terrence Martell, 2005, Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders, *Journal of Financial Markets*, 8, 289-309.

Bae, Kee-Hong, Hasung Jang, and Kyung Suh Park, 2003, Traders' choice between limit and market orders: evidence from the NYSE stocks, *Journal of Financial Markets*, 6, 517-538.

Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *The Journal of Finance,* 50, 1655-1689.

Bloomfield, Robert, Maureen O'Hara and Gideon Saar, 2005, The make or take decision in an electronic market: Evidence on the evolution of liquidity, *Journal of Financial Economics*, 75, 165–199.

Cao, Charles, Oliver Hansch, and Xiaoxin Wang, 2003, The informational content of an open limit order book, *Working Paper*, Pennsylvania State University.

Coppejans, Mark, Ian Domowitz, and Ananth Madhavan, 2004, Liquidity in an automated auction, *Working Paper*, Duke University.

Escribano, Álvaro, and Roberto Pascual, 2006, Asymmetries in bid and ask responses to innovations in the trading process, *Empirical Economics*, 30, 913-946

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets*, 99-134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, Limit order book as a market for liquidity, *Review of Financial Studies*¸ 18, 1171-1217.

Foucault, Thierry, Sophie Moinas, and Erik Theissen, 2006, Does anonymity matter in electronic limit order markets? *Working Paper*, HEC School of Management, Paris.

Franke, Günter, and Dieter Hess, 2000, Information diffusion in electronic and floor trading, *Journal of Empirical Finance*, 7, 455-478.

Greene, William H., 2003, <u>Econometric Analysis</u>, Prentice Hall.

Gouriéroux, C., Monfort, A. and A. Trognon, 1984, Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, 52, 681-700.

Handa, Puneet, and Robert A. Schwartz, 1996, Limit order trading, *The Journal of Finance* 51, 1835-1861.

Handa, Puneet, Robert A. Schwartz, and Ashish Tiwari, 2004, Quote setting and price formation in an order driven market, *Journal of Financial Markets,* forthcoming.

Harvey, A.C., 1992, <u>Forecasting, structural time series models and the Kalman Filter</u>, Cambridge.

Harvey, A.C., E. Ruiz, E. Sentana, 1992, Unobserved component time series models with ARCH disturbances, *Journal of Econometrics*, 52, 129-157

Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance*, 46, 179-207.

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

Hasbrouck, Joel, 1996, Modeling market microstructure time series, in G.S. Maddala and C.R. Rao, eds.: *Handbook of Statistics*, vol. 14, *Statistical Methods in Finance* Elsevier, Holland, Amsterdam.

Hasbrouck, Joel, 1999, The dynamics of discrete bid and ask quotes, *Journal of Finance*, 6, 2109-2142.

Huang, Roger D., and Hans R. Stoll, 1997, The components of the bid-ask spread: a general approach, *Review of Financial Studies*, 10, 995-1034.

Irvine, Paul, Benston, George, and Eugene Kandel, 2000, Liquidity beyond the inside spread: measuring and using information in the limit order book, *Working Paper*, Emory University.

Kaniel, Ron, and Hong Liu, 2001, So what orders do informed traders use?, *Working Paper*, University of Texas at Austin.

Kim, Albert S., 1985, Continuos auctions and insider trading, *Econometrica*, 53, 1315-1335.

Kim, C.J. and C. R. Nelson, 1999, State-Space Models with Regime Switching, MIT press.

Lee, Charles M., Belinda Mucklow, and Mark J. Ready, 1993, Spreads, depths, and the impact of earnings information: an intraday analysis, *Review of Financial Studies*, 6, 345-374.

Lo, Andrew W., A. Craig MacKinlay, and June Zhang, 2002, Econometric models of limit-order executions, *Journal of Financial Economics*, 65, 31-71.

Madhavan, Ananth, 2000, Market microstructure: a survey, *Journal of Financial Markets*, 3, 205-258.

O'Hara, Maureen, 1995, Market Microstructure Theory, Blackwell, Cambridge.

Pardo, Angel, and Roberto Pascual, 2004, On the hidden side of liquidity, *Working Paper Series 2004,* FIN-04-004, Leonard N. Stern School of Business, New York University.

Parlour, Christine A., 1998, Price dynamics and limit order markets, *Review of Financial Studies*, 789-816.

Pascual, Roberto, Bartolomé Pascual-Fuster, and Francisco J. Climent, 2006, Cross-listing, price discovery, and the informativeness of the trading process, *Journal of Financial Markets*, forthcoming.

Pascual, Roberto, and David Veredas, 2004, What pieces of limit order book information do are informative?, *Working Paper Series* 2004, FIN-04-003, Leonard N. Stern School of Business, New York University.

Ranaldo, Angelo, 2004, Order aggressiveness in limit order book markets, *Journal of Financial Markets*, 7, 53-74.

Rindi, Barbara, 2003, Informed traders as liquidity providers: transparency, liquidity and price formation, *Working Paper*, Bocconi University.

Seppi, Duane, 1997, Liquidity provision with limit orders and a strategic specialist, *Review of Financial Studies*, 10, 103–150.

Stoll, Hans, R., 2000, Friction, *Journal of Finance*, 55, 4, 1479-1514.

World Federation of Exchanges, 2004, Annual Report and Statistics, WFE.

TABLE I
Sample statistics

This table provides daily statistics on the 33 Spanish stocks in the sample. We use the weight of each stock in the IBEX-35 index as the proxy for market capitalization. We compute the median of the following daily statistics for each stock: the relative spread is the average ratio of the quoted bid-ask spread over the quote midpoint, weighted by time; depth is measured as the € value of the accumulated volume offered at the 5 best ask and bid quotes, weighted by time; volume traded is measured as the number of shares transacted; finally, order flow is the total number of orders submitted (market, limit, market-to-limit, and cancellations). We compute daily medians for each stock, and we provide cross-sectional statistics of those medians: median, percentiles (25% and 75%) maximum and minimum. We include separated statistics for the 5 largest and the 15 smallest stocks in terms of market capitalization.

Sample	Capitalization (index weight)	Relative Spread	Depth (€)	Volume (shares)	Trades	Order flow
Median	0.75	0.0034	124322.02	347205	365.00	833.50
pctl.(75%)	1.94	0.0039	195705.07	1093215	669.50	1374.50
pctl.(25%)	0.59	0.0024	102173.16	161382	226.50	584.00
Max.	24.68	0.0066	1264221.90	22195189	5675.50	10357.50
Min.	0.18	0.0007	72380.74	95755	86.50	243.50
5 largest stocks						
Median	16.28	0.0011	880163.79	7108202.5	1904.50	3470.00
Over the sample	70.76%			72.94%	48.91%	43.55%
15 smallest stocks						
Median	0.52	0.0038	102173.16	198705	241.00	586.00
Over the sample	7.53%			8.90%	18.02%	21.48%

TABLE II
Estimated co-integration state space model for ask and bid quotes

This table summarizes the estimation of the co-integration state space model for ask and bid quotes in equations [2] to [8] in the paper. The parameters β, ϕ_a, ϕ_b correspond to half the average bid-ask spread, and the persistency of the transitory components of ask and bid quotes, as shown in [T1] and [T2], where $S_{a,t}$ and $S_{b,t}$ are the transitory component of ask and bid quotes, and m_t is the efficient price.

$$\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \beta \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} S_{at} \\ S_{bt} \\ m_t \end{pmatrix} \quad [T1]$$

$$\begin{pmatrix} S_{a,t} \\ S_{b,t} \\ m_t \end{pmatrix} = \begin{pmatrix} \phi_a & \phi_{ab} & 0 \\ \phi_{ba} & \phi_b & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} S_{a,t-1} \\ S_{b,t-1} \\ m_{t-1} \end{pmatrix} + \begin{pmatrix} g_t \\ -g_t \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \end{pmatrix} \quad [T2]$$

The parameters σ_a^2 and σ_b^2 correspond to the average transitory volatility of ask and bid quotes, represented in equation [T3], and α_0, α_1, α_2, and α_3 are the parameters of the EGARGH model in [T4] that represents the volatility of the efficient price.

$$\sigma_{i,t}^2 = \exp\left(\sigma_i^2 + \sum_{j=1}^J \phi_j^i D_{t,j} \right), \ i = \{a,b\} \quad [T3]$$

$$\sigma_{m,t}^2 = \exp\left(\alpha_0 + \alpha_1 \ln\left(\sigma_{m,t-1}^2\right) + \alpha_2 \xi_{m,t-1} + \alpha_3 \left[\left| \xi_{m,t-1} \right| - \sqrt{2/\pi} \right] + \sum_{j=1}^J \phi_j^m D_{t,j} \right) \quad [T4]$$

The model is estimated with the Kalman filter. Panel A provides cross-sectional medians on the estimated parameters for the 33 Spanish stocks in our sample. Panel B provides cross-sectional correlations among the parameters and the following statistics: the relative spread is the average ratio of the quoted bid-ask spread over the quote midpoint, weighted by time; depth is measured as the € value of the accumulated volume offered at the 5 best ask and bid quotes, weighted by time; volume traded is measured as the number of shares transacted; finally, order flow is the total number of orders submitted (market, limit, market-to-limit, and cancellations). Panel C provides cross-sectional correlations between the same statistics than in Panel B with the estimated short-term component of ask and bid quotes in [T1] and with the transitory and long run volatilities in [T3] and [T4].

Panel A: Estimated coefficients

Sample	β	ϕ_a	ϕ_b	ϕ_{ab}	ϕ_{ba}	σ_a	σ_b	α_0	α_1	α_2	α_3
Median	0.0207	0.7890	0.6269	0.2592	0.2174	0.0002	0.0002	-0.4000	0.9402	0.0093	0.3102
pctl.(75%)	0.0356	0.8285	0.7006	0.3455	0.3200	0.0005	0.0004	-0.2696	0.9520	0.0384	0.3625
pctl.(25%)	0.0125	0.7051	0.5589	0.2087	0.1829	0.0001	0.0001	-0.6203	0.8972	-0.0080	0.2287
Max.	0.0663	0.9862	0.8413	0.9900	0.5771	0.0015	0.0013	-0.0788	0.9820	0.0608	0.4838
Min.	0.0062	0.4049	0.3487	0.1153	-0.2135	0.0000	0.0000	-1.3404	0.8259	-0.0288	0.1321
5 largest stocks											
Median	0.0080	0.6069	0.4756	0.3782	0.4218	0.0000	0.0001	-0.1425	0.9759	-0.0145	0.1578
15 smallest stocks											
Median	0.0233	0.8174	0.6773	0.2105	0.2032	0.0002	0.0002	-0.5626	0.8972	0.0381	0.3625

Panel B: Correlations with estimated parameters

Statistic	β	ϕ_a	ϕ_b	ϕ_{ab}	ϕ_{ba}	σ_a	σ_b	α_0	α_1	α_2	α_3
Capitalization	-0.4102	-0.4542	-0.6627	0.5505	0.7525	-0.3110	-0.2893	0.4384	0.4816	-0.4188	-0.5775
Relative spread	0.6686	0.7314	0.7765	-0.8445	-0.8437	0.4332	0.4285	-0.4507	-0.5475	0.6576	0.7149
Depth (€)	-0.4574	-0.4936	-0.6945	0.6026	0.7652	-0.3569	-0.3313	0.4818	0.5388	-0.4876	-0.6388
Volume	-0.3949	-0.3991	-0.6517	0.4586	0.7436	-0.3029	-0.2795	0.3829	0.4274	-0.3747	-0.5312
Trades	-0.3852	-0.5761	-0.6565	0.6184	0.8385	-0.2530	-0.2384	0.4576	0.4981	-0.4041	-0.5882
Order flow	-0.3498	-0.5795	-0.6299	0.6145	0.8258	-0.2127	-0.2028	0.4632	0.4936	-0.3931	-0.5714

Panel C: Correlations with estimated factors and volatilities

Statistic	σ_a^2	σ_b^2	σ_m^2	S_a	S_b
Capitalization	-0.3143	-0.2851	-0.0023	-0.2482	-0.3954
Relative spread	0.4980	0.3811	-0.2127	0.3745	0.5392
Depth (€)	-0.3538	-0.3222	-0.0269	-0.2801	-0.4399
Volume	-0.3014	-0.2768	0.0259	-0.2304	-0.3872
Trades	-0.2901	-0.2596	0.4110	-0.2771	-0.2772
Order flow	-0.2588	-0.2294	0.4856	-0.2666	-0.2272

TABLE III
VAR model for limit order book variables

Panel A summarizes the estimation of the following VAR model by OLS,

$$y_t = \gamma_0 + \sum_{j=1}^{3} \gamma_j ' y_{t-j} + w_t .$$ [T5]

where y_t is the vector LOB variables. This vector includes: the relative bid-ask spread (S_t); the displayed depth at the best ask (D_t^a) and bid (D_t^b) quotes; the accumulated depth at up to four ticks from the quote midpoint ($BD_t^a(4)$, $BD_t^b(4)$); the order imbalance on the first level of the LOB (AsD_t); the order imbalance beyond the best quotes ($AsBD_t(4)$), and the bi-dimensional liquidity measure in [1] for a trade of size equal to 2 times the NMS ($BLM_t(2)$). For each explanatory variable we provide the sum of the lagged coefficients that are statistically significant (at least) at the 5% level. Panel B provides the residual cross-equation correlations, computed from the estimated w_t in [T5].

Panel A: Coefficients

	Dependent variable							
	S	D^a	D^b	AsD	$BLM(2)$	$BD^a(4)$	$BD^b(4)$	$AsDB(4)$
S	0.6678	-0.0407	-0.0302	-0.0338	-0.0469	-0.0165	-0.0207	____
D^a	-0.0013	0.6891	0.0469	0.0993	0.0058	0.0051	-0.0063	-0.0198
D^b	-0.0017	0.0229	0.6368	0.0596	0.0070	0.0025	0.0161	-0.0033
AsD	0.0024	-0.0051	-0.0047	0.5477	-0.0077	0.0073	0.0164	0.0227
$BLM(2)$	0.1462	0.0297	0.0411	-0.0032	0.8555	-0.0721	-0.0395	-0.0896
$BD^a(4)$	-0.0135	0.0616	0.0004	____	-0.0145	0.8169	0.0648	0.0937
$BD^b(4)$	-0.0154	0.0120	0.0748	0.0178	-0.0143	0.0672	0.7950	0.0673
$AsDB(4)$	0.0128	-0.0193	-0.0041	0.0444	0.0120	-0.0658	-0.0668	0.6248
R^2	0.6249	0.4959	0.4119	0.3922	0.6536	0.6166	0.5659	0.4974
Obs.	366920							

Panel B: Correlation matrix of residuals

	S	D^a	D^b	AsD	$BLM(2)$	$BD^a(4)$	$BD^b(4)$	$AsDB(4)$
S	1.0000							
D^a	0.0738	1.0000						
D^b	0.0787	0.0287	1.0000					
AsD	-0.0005	0.5851	0.5693	1.0000				
$BLM(2)$	0.8314	-0.0101	0.0052	-0.0231	1.0000			
$BD^a(4)$	-0.1354	0.3410	0.0051	0.1728	-0.1412	1.0000		
$BD^b(4)$	-0.1359	0.0092	0.3585	0.1859	-0.1329	0.1066	1.0000	
$AsDB(4)$	-0.1504	0.1526	0.1735	0.2850	-0.1194	0.5973	0.6141	1.0000

TABLE IV
Estimation of the SVAR model: 5-minute periodicity

This table summarizes the estimation of the following structural VAR (SVAR) model for the 33 stocks in our sample,

$$Az_t = A_0 + \sum_{j=1}^{p} A_j z_{t-1} + F'x_t + v_t \,. \tag{T6}$$

The "Best quotes" model combines a long run volatility proxy with the relative bid-ask spread (S), the displayed depth at the best ask (D^a) and bid (D^b) quotes, and the order imbalance on the first level of the LOB (AsD). The "Book" model combines the informational volatility measure with the accumulated depth at up to four ticks from the quote midpoint (BD^a, BD^b); the order imbalance beyond the best quotes ($AsDB$), and the bi-dimensional liquidity measure in equation [1] in the paper (BLM), for a trade of size equal to 2 times the NMS. Finally, the "Best + book" includes all the variables in the two previous models. All time series are computed over 5-minute intervals, and averaged weighting by time. The vector of control variables (x_t) includes the number of trades and the orders submitted (market, limit, market-to-limit orders, and cancellations). The deterministic intraday component of each time series has been removed in a previous step. The lag length of the model is fixed at one. The matrix of contemporaneous relationships (A) allows contemporaneous causality running from volatility to the LOB variables, but no causality running from the LOB variables to long run volatility. Moreover, the variance-covariance matrix of (v_t) allows for contemporaneous correlation among the innovation of the LOB variables, but imposes no correlation between the innovations in long run volatility and the innovations in the LOB. With these restrictions the model is exactly identified. Panel A shows the free parameters of matrix A. Panel B provides the estimated coefficients of the reduced form of model [T6], a VAR model. Only the results for the equation of the long run volatility are reported. We also report Granger causality tests for each of the LOB variables. Finally, "***", "**", and "*" means statistically significant at the 1%, 5% and 10% level, respectively.

Panel A: Contemporaneous coeficients of the SVAR model

Explanatory variable	Model Best quotes	Book	Best+Book
S	-27.2774 ***		-28.2788 ***
D^a	4.8156 *		4.3894
D^b	1.8513		1.1206
AsD	-14.6514 ***		-21.2577 ***
$BLM(2)$		-26.1284 ***	-26.1910 ***
$BD^a(4)$		29.0610 ***	28.8908 ***
$BD^b(4)$		31.9487 ***	31.7740 ***
$AsDB(4)$		-14.5243 ***	-14.5669 ***

Panel B: Coeficients of the reduced form VAR (x10000 except σ_m)

Explanatory variable	Model Best quotes	Book	Best+Book
σ_m	0.9413 ***	0.9414 ***	0.9413 ***
S	0.5970 ***		-0.5570 ***
D^a	-0.0161		0.2140 ***
D^b	0.0053		0.2180 ***
AsD	0.0061		-0.2550 ***
$BLM(2)$		0.9830 ***	1.4680 ***
$BD^a(4)$		0.0267 **	-0.0460 ***
$BD^b(4)$		0.0321 ***	-0.0461 ***
$AsDB(4)$		-0.0192	0.0590 ***
T	0.0364 ***	0.0387 ***	0.0285 ***
O	0.5980 ***	0.5760 ***	0.5750 ***
R^2	0.9337	0.9342	0.9342
No. of obs.	374131	373854	373865
Causality test:			
S	839.040 ***		216.670 ***
D^a	2.915		212.140 ***
D^b	0.305		231.600 ***
AsD	0.682		318.730 ***
$BLM(2)$		1814.800 ***	1247.400 ***
$BD^a(4)$		5.026 **	12.383 ***
$BD^b(4)$		7.296 ***	12.531 ***
$AsDB(4)$		2.472	19.979 ***
ALL	865.100 ***	1839.900 ***	2290.700 ***

TABLE V
Estimation of the SVAR model: Market capitalization

This table summarizes the estimation of the following structural VAR (SVAR) model for the 5 largest and 15 smallest stocks in our sample by market capitalization, with time series computed over 5-minute intervals,

$$Az_t = A_0 + \sum_{j=1}^{p} A_j z_{t-1} + F' x_t + v_t .$$ [T7]

The "Best quotes" model, the "Book" model, and the "Best + book" model are as described in Table IV. The variables are: a proxy for the long run volatility, estimated using co-integration state-space model for ask and bid quotes, the relative bid-ask spread (S), the displayed depth at the best ask (D^a) and bid (D^b) quotes, the order imbalance on the first level of the LOB (AsD), the accumulated depth at up to four ticks from the quote midpoint (BD^a, BD^b), the order imbalance beyond the best quotes ($AsDB$), and the bi-dimensional liquidity measure in equation [1] in the paper (BLM), for a trade of size equal to 2 times the NMS. All LOB variables are averaged weighting by time. The control variables in x_t are the number of trades and the number of orders submitted. The deterministic intraday component of each time series has been previously removed. The lag length is fixed at one. The matrix of contemporaneous relationships (A) allows causality running from volatility to the LOB variables, but not the other way around. The variance-covariance matrix of (v_t) allows for contemporaneous correlation among the innovation of the LOB variables, but imposes no correlation between the innovations in long run volatility and the innovations in the LOB. The model is exactly identified. Panel A reports the estimated free parameters of matrix A. Panel B provides the estimated coefficients of the reduced form (VAR) model. Only the results for the equation of the long run volatility are reported. Granger causality tests for each of the LOB variables are also reported. Finally, "***", "**", and "*" means statistically significant at the 1%, 5% and 10% level, respectively.

Panel A: Contemporaneous coeficients of the SVAR model

Explanatory variable	5 largest stocks			15 smallest stocks		
	Model			Model		
	Best quotes	Book	Best+Book	Best quotes	Book	Best+Book
S	-6.9129 ***		-7.6022 ***	-30.1621 ***		-30.8687 ***
D^a	10.1605		10.1285	-2.5712		-3.0289
D^b	7.5044		7.4141	-7.0201		-7.5474
AsD	0.5991		-2.8512	-14.4934 **		-29.7451 ***
$BLM(2)$		-10.3284 ***	-10.4442 ***		-25.8092 ***	-25.8508 ***
$BD^a(4)$		28.4337 ***	28.3287 ***		18.6926 ***	18.4849 ***
$BD^b(4)$		20.0839 ***	19.9144 ***		29.7383 ***	29.5356 ***
$AsDB(4)$		1.1082	1.0863		-14.2947 **	-14.3714 **

Panel B: Coeficients of the reduced form VAR (x10000 except σ_m)

Explanatory variable	5 largest stocks			15 smallest stocks		
	Model			Model		
	Best quotes	Book	Best+Book	Best quotes	Book	Best+Book
σ_m	0.9413 ***	0.9086 ***	0.9087 ***	0.9413 ***	0.9434 ***	0.9433 ***
S	0.7000 ***		-0.5100 ***	0.7400 ***		-0.5810 ***
D^a	-0.0434		0.2180 ***	-0.0118		0.2210 ***
D^b	0.0317		0.2810 ***	-0.0017		0.2010 ***
AsD	-0.0187		-0.2600 ***	0.0085		-0.2640 ***
$BLM(2)$		0.9360 ***	1.3610 ***		1.2340 ***	1.7470 ***
$BD^a(4)$		-0.1880 ***	-0.2680 ***		0.0762 ***	0.0015
$BD^b(4)$		-0.1180 ***	-0.2270 ***		0.0952 ***	0.0228
$AsDB(4)$		0.1370 ***	0.2250 ***		-0.0744 ***	0.0066
T	0.2190 ***	0.2470 ***	0.2310 ***	0.0285 **	0.0305 **	0.0197
O	0.7670 ***	0.7360 ***	0.7300 ***	0.6150 ***	0.5890 ***	0.5890 ***
R^2	0.8677	0.8678	0.8679	0.9447	0.9449	0.9450
No. of obs.	56679	56639	56639	170133	169989	169989
Causality test:						
S	91.107 ***		12.310 ***	512.860 ***		95.543 ***
D^a	2.248		18.991 ***	0.607		92.849 ***
D^b	0.865		30.674 ***	0.014		86.807 ***
AsD	0.390		26.732 ***	0.661		146.930 ***
$BLM(2)$		145.720 ***	81.352 ***		1090.700 ***	686.920 ***
$BD^a(4)$		16.624 ***	28.868 ***		18.341 ***	0.005
$BD^b(4)$		6.797 ***	20.825 ***		29.168 ***	1.390
$AsDB(4)$		8.368 ***	19.718 ***		16.659 ***	0.113
ALL	99.376 ***	182.000 ***	220.150 ***	523.730 ***	1099.000 ***	1305.300 ***

TABLE VI
Estimation of the SVAR model: 1-minute periodicity

This table summarizes the estimation of the structural VAR (SVAR) model [T8] for the 33 stocks in our sample with time series computed over 1-minute intervals,

$$Az_t = A_0 + \sum_{j=1}^{p} A_j z_{t-1} + F'x_t + v_t .$$ [T8]

The "Best quotes" model, the "Book" model, and the "Best + book" model are as described in Tables IV and V. The variables in the model are a proxy for the long run volatility, estimated using a co-integration state-space model for ask and bid quotes; the relative bid-ask spread (S); the displayed depth at the best ask (D^a) and bid (D^b) quotes; the order imbalance on the first level of the LOB (AsD); the accumulated depth at up to four ticks from the quote midpoint (BD^a, BD^b); the order imbalance beyond the best quotes ($AsDB$), and the bi-dimensional liquidity measure in equation [1] in the paper (BLM), for a trade of size equal to 2 times the NMS. All the LOB time series are averaged weighting by time. The vector of control variables (x_t) includes the number of trades and the number of orders submitted. The deterministic intraday component of each time series has been previously removed. The lag length of the model is fixed at one. The matrix of contemporaneous relationships (A) allows causality running from volatility to the LOB variables, but not the other way around. The variance-covariance matrix of (v_t) allows for contemporaneous correlation among the innovation of the LOB variables, but imposes no correlation between the innovations in long run volatility and the innovations in the LOB. The model is exactly identified. Panel A reports the estimated pooled free parameters of matrix A. Panel B provides the estimated pooled coefficients of the reduced form (VAR) model. Only the results for the equation of the long run volatility are reported. Granger causality tests for each of the LOB variables are also reported. Finally, "***", "**", and "*" means statistically significant at the 1%, 5% and 10% level, respectively.

Panel A: Contemporaneous coeficients of the SVAR model

Explanatory variable	Model Best quotes	Book	Best+Book
S	-24.0934 ***		-24.1677 ***
D^a	2.0479 **		1.9184 *
D^b	8.2719 ***		8.1579 ***
AsD	-2.3084 *		1.6370
$BLM(2)$		-20.4372 ***	-20.4505 ***
$BD^a(4)$		8.1462 ***	8.1556 ***
$BD^b(4)$		7.8460 ***	7.8392 ***
$AsDB(4)$		-2.3334 *	-2.3039 *

Panel B: Coeficients of the reduced form VAR (x10000 except σ_m)

| Explanatory variable	Model Best quotes	Book	Best+Book
σ_m	0.8739 ***	0.8740 ***	0.8740 ***
S	0.5690 ***		-0.2370 ***
D^a	0.0057		0.1520 ***
D^b	0.0450 ***		0.1970 ***
AsD	-0.0313 ***		-0.1610 ***
$BLM(2)$		0.9370 ***	1.1430 ***
$BD^a(4)$		0.0674 ***	0.0159 *
$BD^b(4)$		0.0589 ***	-0.0147 *
$AsDB(4)$		-0.0817 ***	-0.0280 ***
T	0.0452 ***	0.0457 ***	0.0436 ***
O	0.2090 ***	0.2030 ***	0.2030 ***
R^2	0.7731	0.7732	0.7733
No. of obs.	1905913	1905527	1905527
Causality test:			
S	1945.700 ***		131.090 ***
D^a	1.179		344.540 ***
D^b	89.951 ***		615.930 ***
AsD	48.534 ***		427.640 ***
$BLM(2)$		4042.400 ***	2424.200 ***
$BD^a(4)$		67.935 ***	3.326 *
$BD^b(4)$		52.492 ***	2.862 *
$AsDB(4)$		102.510 ***	10.860 ***
ALL	2251.000 ***	4167.500 ***	4835.600 ***

FIGURE 1
Estimated average bid-ask spread

This figure is a scatter-plot of the estimated average bid-ask spread using the co-integration state-space model for ask and bid quotes in [2]-[8] in the paper, given by 2β in [2], and the median bid-ask spread observed for the 33 Spanish stocks in the sample.
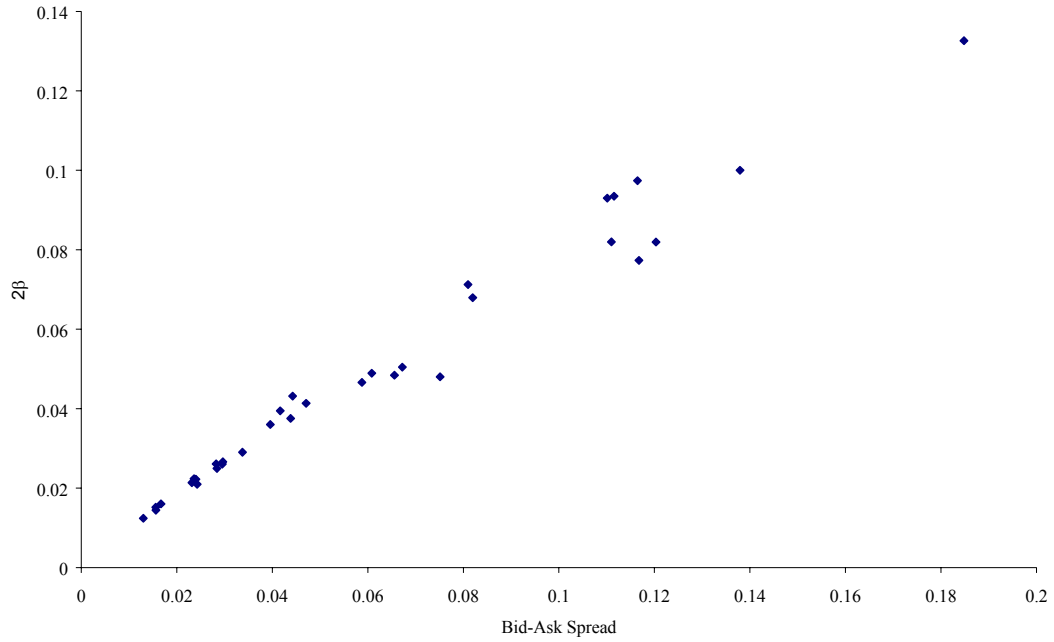
FIGURE 2
## Estimated intraday regularities in the bid-ask spread

This figure shows the cross-sectional median of the estimated intraday regular patterns of the transitory components of ask and bid quotes for a sample of 33 stocks listed in the electronic platform of the SSE. Transitory components of quotes are characterized in equations [2] and [3] of the co-integration state-space model given by equations [2] to [8] in the paper. The unobserved components of ask and bid quotes are estimated using the Kalman filter. The regular patterns in the transitory components of ask and bid quotes are given by the parameters ϕ_j in equation [4]. We compare the estimated regularities ($2\hat{\phi}_j$) with the observed cross-sectional median level of the bid-ask spread during the same time intervals. We split the trading session of the SSE in the following intervals: [9:00, 9:30), [9:30, 10:00), [10:00, 13:00), [13:00, 15:00), [15:00, 15:30), [15:30, 16:00), [16:00, 16:30), [16:30, 17:00), [17:00, 17:30]. The figure reports the deviation of each coefficient or median with respect to the control interval [13:00, 15:00).

40
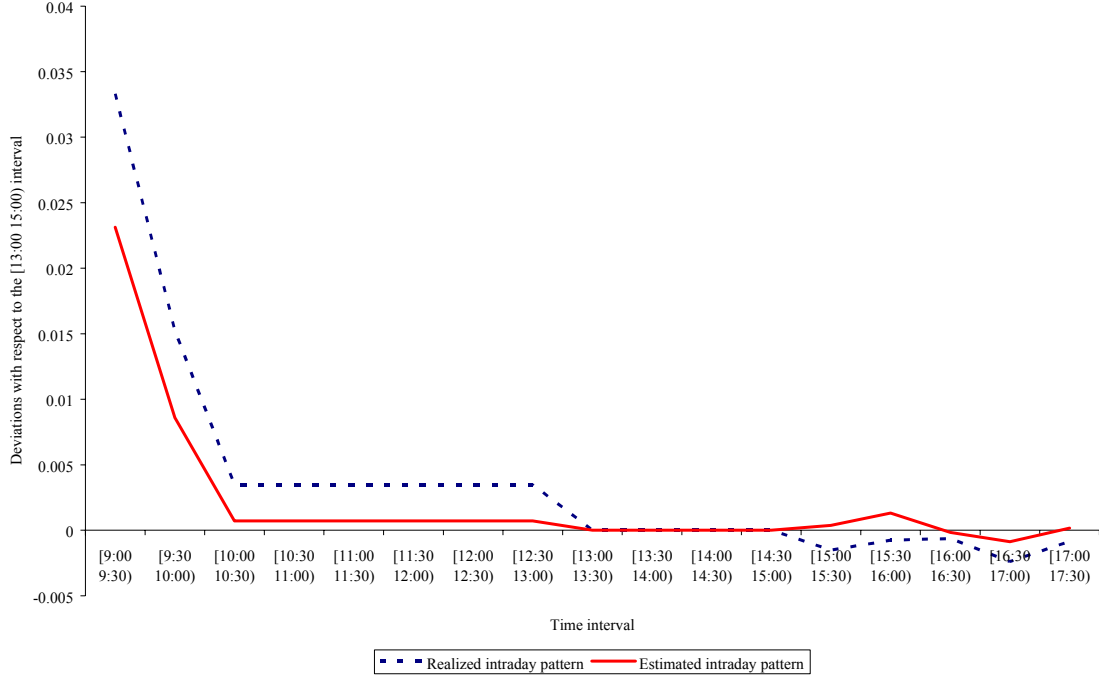
FIGURE 3
Estimated intraday regularities in volatility

This figure reports the estimated intraday deterministic patterns in the transitory and informational components of the volatility of ask and bid quotes for the 33 Spanish stocks in our sample. Volatility is decomposed in its theoretical components using the co-integration state-space model given by equations [2] to [8] in the paper. Transitory volatility is described by the deterministic process in equation [7]. Informational volatility is modeled as the EGARCH model in equation [8]. The estimated deterministic patterns are given by the parameters ϕ_j^a and ϕ_j^b in [7] for the transitory volatility, and ϕ_j^m in [8] for the informational volatility.

We split the trading session of the SSE in the following intervals: [9:00, 9:30), [9:30, 10:00), [10:00, 13:00), [13:00, 15:00), [15:00, 15:30), [15:30, 16:00), [16:00, 16:30), [16:30, 17:00), [17:00, 17:30). The table reports the cross-sectional median of the each parameter, but expressed in deviations with respect to the control interval [13:00, 15:00).
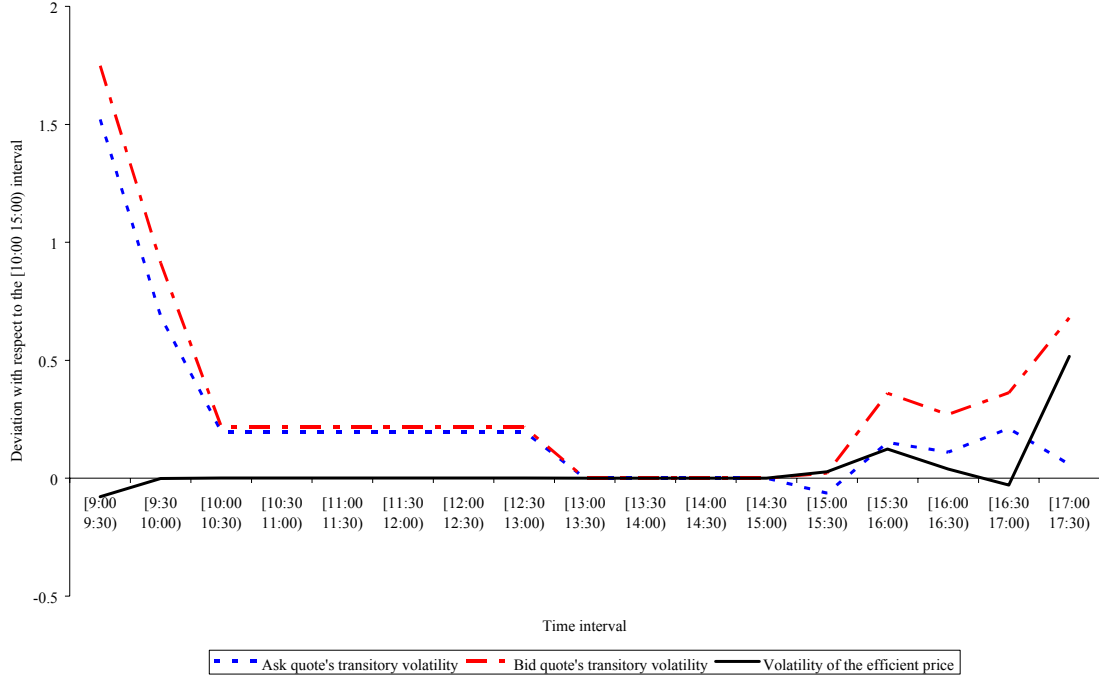
Appendix

In this appendix, we provide some details about how do we adjust the state space model in [2]-[8] so as to apply the Kalman filter. We rely in Harvey, Ruiz, and Sentana (1992) and Kim and Nelson (1999, chapter 6). First, notice that [2] and [3] are the so-called measurement and transition equations of the state space model. The EGARCH model in [8] is defined on the innovation of the efficient price in the transition equation, which can be rewritten as

$$
\begin{pmatrix} S_{a,t} \\ S_{b,t} \\ m_t \end{pmatrix} = \begin{pmatrix} \phi_a & \phi_{ba} & 0 \\ \phi_{ab} & \phi_b & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} S_{a,t-1} \\ S_{b,t-1} \\ m_{t-1} \end{pmatrix} + \begin{pmatrix} g_t \\ -g_t \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{a,t}^* \\ \varepsilon_{b,t}^* \\ \varepsilon_{m,t}^* \end{pmatrix},
\qquad [3\text{'}]
$$

where

$$
\begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{a,t}^* \\ \varepsilon_{b,t}^* \\ \varepsilon_{m,t}^* \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Omega_t \right),
\qquad [5\text{'}]
$$

and

$$
\Omega_t = \begin{pmatrix} \sigma_{a,t}^2 & 0 & 0 \\ 0 & \sigma_{b,t}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{m,t}^{*2} \end{pmatrix},
\qquad [6\text{'}]
$$

with $\sigma_{m,t}^{*2}$ following the EGARCH model [8]. This model cannot be estimated by the standard Kalman filter since we would need to know $\sigma_{m,t}^{*2}$, which would appear in the prediction equations of the filter. The issue is that the EGARCH model depends on past unobserved shocks, and even the knowledge of past factors does not imply the knowledge of past unobserved shocks. The problem is solved replacing past shocks and past absolute values of the shocks in [8] by its conditional expectations,

$$
\ln \sigma_{m,t}^{*2} = \alpha_0 + \alpha_1 \ln \left(\sigma_{m,t-1}^{*2} \right) + \alpha_2 E\left[\xi_{m,t-1}^* \right] + \alpha_3 \left[E\left[\left| \xi_{m,t-1}^* \right| \right] - \sqrt{2/\pi} \right] + \sum_{j=1}^{J} \phi_j^m D_{t,j}, \quad [8\text{*}]
$$

where $\xi_{m,t}^* = \varepsilon_{m,t}^* / \sigma_{m,t}^*$ is the standardized shock.

Therefore, [8*] represents an approximation to the real process [8] and, hence, the quasi-optimal term proposed by Harvey *et al* (1992) applies. To obtain the conditional expectations the state is augmented yielding new measurement and transition equations,

$$
\begin{pmatrix} a_t \\ b_t \end{pmatrix} = \beta \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} S_{a,t} \\ S_{b,t} \\ m_t \\ \varepsilon_{m,t}^* \end{pmatrix} \qquad [2^*]
$$

and

$$
\begin{pmatrix} S_{a,t} \\ S_{b,t} \\ m_t \\ \varepsilon_{m,t}^* \end{pmatrix} = \begin{pmatrix} \phi_a & \phi_{ba} & 0 & 0 \\ \phi_{ab} & \phi_b & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} S_{a,t-1} \\ S_{b,t-1} \\ m_{t-1} \\ \varepsilon_{m,t-1}^* \end{pmatrix} + \begin{pmatrix} g_t \\ -g_t \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \\ \varepsilon_{m,t}^* \end{pmatrix}, \qquad [3^*]
$$

where

$$
\begin{pmatrix} \varepsilon_{a,t} \\ \varepsilon_{b,t} \\ \varepsilon_{m,t} \\ \varepsilon_{m,t}^* \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Omega_t \right) \qquad [5^*]
$$

and

$$
\Omega_t = \begin{pmatrix} \sigma_{a,t}^2 & 0 & 0 & 0 \\ 0 & \sigma_{b,t}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{m,t}^{*2} \end{pmatrix}. \qquad [6^*]
$$

The restriction $\sigma_{m,t}^2 = 0$ in the diagonal of Ω_t is necessary to identify $\sigma_{m,t}^{*2}$ and α_0. The model given by equations [2*]-[3*]-[4]-[5*]-[6*]-[7]-[8*] can be estimated by standard Kalman filter.