Estimating the Price Impact of Trades in a High-Frequency Microstructure Model with Jumps *

Eric Jondeau[a], Jérôme Lahaye[b], and Michael Rockinger[c]

September 2015

Abstract

We estimate a general microstructure model of the transitory and permanent impact of order flow on stock prices. Jumps are detected in both the transaction price (observation equation) and fundamental value (state equation). The model's parameters and variances are updated in real time. Prices can be altered by both the size and direction of trades, and the effects of buy-initiated and sell-initiated trades are different. We estimate this model using tick-by-tick data for 12 large-capitalization stocks traded on the Euronext-Paris Bourse. We find that, at tick frequency, the overnight return, the intraday jumps, and the continuous innovations represent approximately 7%, 8.5%, and 36.7% of the total variation of stock returns. The microstructure model explains on average 47.7% of the total variation. Once jumps are filtered and parameters are estimated in real time, we also find that the price impact of trades is symmetric on average. However, the price of highly liquid stocks with a large proportion of sell-initiated orders tends to be more sensitive to buy trades, whereas the price of less liquid stocks with a large proportion of buy-initiated orders tends to be more sensitive to sell trades.

Keywords: Microstructure model, jumps, noise, volatility, Kalman filter, particle filter.

JEL classification: C10, C14, C22, C41, C51, G1

[a]Swiss Finance Institute and University of Lausanne, Switzerland. E-mail: eric.jondeau@unil.ch.

[b]Fordham University, New York, USA. E-mail: jlahaye@fordham.edu.

[c]Swiss Finance Institute and University of Lausanne, Switzerland. E-mail: michael.rockinger@unil.ch.

*Neither do the Swiss Finance Institute, nor the University of Lausanne necessarily endorses the views expressed in this paper. The 3rd author is grateful to CREST for its hospitality. We are grateful to Hedibert Lopes and Ruey Tsay for making available particle filter codes that helped expedite the development of our own codes. We are grateful for helpful comments from Gaëlle Le Fol, Serge Darolles, Siem Jan Koopman, seminar participants at National University Singapore, CREST, Paris Sorbonne, as well as participants at the 11th Oxmetrics Conference, the 1st Conference on Recent Developments in Financial Econometrics and Applications, and two anonymous referees.

1 Introduction

A main focus of the market microstructure literature is the formation of prices of securities traded at high frequency. In most of these models, the difference between the transaction price and the fundamental (or equilibrium) value of a security allows for identification of the main factors driving the price dynamics. The dynamics of the fundamental value is primarily driven by revisions in traders' beliefs, which may have two sources: (1) new public information and (2) changes in order flow (Glosten and Milgrom, 1985, and Glosten and Harris, 1988).

In standard microstructure models (Glosten and Milgrom, 1985, or Kyle, 1985), transaction price departs from the fundamental value because of asymmetric information. Furthermore, transient discrepancies between transaction price and fundamental value reflecting dealer's compensation for order processing or inventory costs (see also Roll, 1984, and Amihud and Mendelsohn, 1986). The residual term then captures the effects of stochastic rounding errors.

The first papers to estimate fully fledged versions of this model using transaction-level data are Brennan and Subrahmanyam (1996) and Madhavan, Richardson, and Roomans (1997, thereafter MRR). Several extensions of the model have been proposed. Sadka (2006) introduces the size of trades as a driver of changes in order flow. Dufour and Engle (2000) and Engle and Sun (2007) introduce the effect of duration between trades of a given stock as an important driver of the volatility of fundamental innovations, confirming the importance of a transaction-level analysis (following theoretical models of Diamond and Verrecchia, 1987, and Easley and O'Hara, 1992). Other contributions estimating market microstructure models in tick time are Rydberg and Shephard (2003), Frijns and Schotman (2004), and Bos (2008).

Although it is essential to understand how transaction prices are determined at the tick level, estimation of these models is plagued by the statistical properties of tick-by-tick data. These properties are the focus of another strand of literature that addresses "microstructure noise" in the measurement of return volatility at high frequency. In this literature, the object of interest is the volatility of the fundamental value. The difference between the observed price and the fundamental value is interpreted as noise caused by microstructure factors such as liquidity or information frictions, tick-size discretization, rounding errors, and so forth (see Bandi and Russell, 2008). The measure of the true variance (integrated variance) is thus contaminated by microstructure noise, so that its estimate (realized variance) diverges from the true variance. Diebold and Strasser (2013) have investigated the magnitude of the bias due to the correlation between the fundamental value and the microstructure noise in a set of structural models. Most statistical papers address this issue by reducing the sampling frequency to mitigate the effect of the noise (Andersen et al., 2001, Barndorff-Nielsen and Shephard, 2002). There is no consensus about the optimal sampling frequency, however, leading Aït-Sahalia, Mykland, and Zhang (2005) to conclude that "modeling the noise and using all the data is a better solution, even if one misspecifies the noise distribution."

An additional difficulty in the measurement of integrated variance is the presence of extreme price changes (or jumps) in the data process, jumps typically caused by large orders. Jumps constitute an additional, probably more severe, source of divergence of realized variance from integrated variance, as the former also incorporates the effects of jumps. In continuous-time models, bipower variation measures provide consistent estimates of integrated variance, even in the presence of stochastic volatility and large but infrequent jumps (Barndorff-Nielsen and Shephard, 2004 and 2006). The detection of jumps has also been investigated in a parametric framework by Johannes, Polson, and Stroud (2009), Bos (2008), and Duan and Fulop (2007), papers that rely on simulation techniques such as Markov Chain Monte Carlo algorithms or particle filters.

To our knowledge, this paper is the first to estimate a complete microstructure model with time-variability of parameters and large but infrequent jumps. We start by extending the model proposed by MRR (1997) and Sadka (2006), and incorporating into it most of the tick data stylized facts. Our model describes the joint dynamics of the transaction price and the fundamental value: changes in beliefs are captured by innovations in order flow but also by innovations in order size (Sadka, 2006). As we use tick data, we can capture the effects of durations between trades on return volatility (Dufour and Engle, 2000). Following a recent contribution of Hameed, Kang, and Viswanathan (2010), sell and buy orders are allowed to have asymmetric effects on prices. Our model also incorporates seasonal patterns of the variables and time-varying parameters and variances. The model parameters, the variances, and the fundamental value are updated in real time. Jumps are identified through an outlier detection procedure, without the need for parametric assumptions about their distribution, frequency, or size. This procedure allows us to detect jumps in real time in both fundamental value (permanent or innovation jumps) and transaction price (transitory or observation jumps) dynamics, following the approach of Maiz, Miguez, and Djuric (2009).

Our estimation strategy relies on the particle filter technique, which estimates the unobservable fundamental value using the observed transaction price. Sadka (2006) uses OLS regressions to estimate model's parameters. MRR estimate the complete model (with fundamental value dynamics) using GMM technique over different intraday trading intervals to allow parameter variability. We do not pursue these estimation approaches because our goal is to estimate the dynamics of the parameters and detect jumps in real time. Unlike the Kalman filter, the particle filter accommodates non-linear models and non-Gaussian innovations, for instance when innovations are perturbed by jumps. The particle filter also allows for sequential or online filtering, so that model parameters and variances can be estimated in real time after detection and deletion of jumps (Kitagawa, 1998).

We illustrate our methodology using 12 of the largest-capitalization stocks traded on the Euronext-Paris Bourse. By focusing on all trades over a two-month period, with numbers of trades ranging between 686 and 4,468 per day, we identify several regularities across time (intradaily seasonality patterns) and across firms (differences in price impact). We find that, at tick frequency, overnight return, the intraday jumps, and the continuous innovations represent approximately 7%, 8.5%, and 36.7% of the total variation of stock returns. The microstructure model explains on average 47.7% of the total variation. In addition, the estimates of the price impact of trades increase by 10 to 30% when the model's parameters are time-varying and jumps are taken into account. We then investigate asymmetries in price impact, depending on whether the trade is buy- or sell-initiated. While previous work on block trades by institutional investors reports that buy trades have greater price impact than sell trades (Holthausen, Leftwich, and Mayers, 1987, Chan and Lakonishok, 1993, Keim and Madhavan, 1995), we find that this asymmetry can be reversed in the case of relatively less liquid firms with a large proportion of buy-initiated orders. This result is consistent with the theoretical analysis of Saar (2001).

The rest of the paper is organized as follows. Section 2 presents the general microstructure model and summarizes our estimation methodology, which is described with more details in the Appendix. Section 3 presents our data set of large-capitalization stocks on Euronext-Paris and provides some preliminary results on the tick-by-tick price dynamics. Section 4 reports our results of the estimation of the microstructural model and analyzes the price impact of trades. Section 5 concludes.

2 A Microstructure Model for Prices

2.1 A General Model

In this section, we describe our general microstructure model of price formation. As the model is formulated in tick time, we denote by t_k the time of the k-th trade on a given day. Observations are randomly spaced through time, so we introduce $\tau_k = t_k - t_{k-1}$, the duration between trades $k-1$ and k. We consider an efficient market in which the fundamental value of a stock, denoted m_k, varies through time according to changes in the beliefs of market participants. We assume that revisions of beliefs are correlated with innovations in order flow (as in Glosten and Milgrom, 1985, and MRR, 1997) and with the innovation in the net order imbalances (as in Glosten and Harris, 1988, and Madhavan and Smidt, 1991). The actual transaction price of a stock, denoted y_k, departs from the fundamental value due to the transitory effects of the incoming trade. We assume that the difference between the actual price and the fundamental value reflects the effects of the order flow and the net order imbalance but that it may also reflect the cost of processing the order as well as rounding errors or stale quotes (Roll, 1984). Before describing the model, we define the two main drivers of the price dynamics. The first driver is the trade direction, denoted by D_k, which takes a value of $+1$ if a trade at time t_k is buy-initiated, i.e., if the trade takes place on the ask side of the order book, and a value of -1 if the trade is sell-initiated. To distinguish potential asymmetries in the dynamics of the order book, we will separately consider the effects of buy-initiated and sell-initiated trades. Consequently, we introduce the dummies I_k^+ (I_k^-), which take a value of 1 if the trade at time t_k is buy (sell)-initiated and 0 otherwise. Clearly, $D_k = I_k^+ - I_k^-$. The second driver is the signed trade size, which captures trade informativeness (Hasbrouk, 1991, Foster and Viswanathan, 1993, Brennan and Subrahmanyam, 1996). We denote the trade size by $V_k = \log(N_k)$, where N_k is the number of shares traded at trade k. The signed trade size is then defined as $DV_k = D_k V_k$. Finally, we also allow fundamental value to depend on durations between trades. Diamond and Verrecchia (1987), Easley and O'Hara (1992), and Parlour (1998) emphasize the theoretical relationship between news and trade frequency, an effect that has been empirically measured in various papers: Dufour and Engle (2000) find that no trade implies no information, while Engle and Sun (2007) show that volatility is related to duration but less than linearly.

The model describing the price dynamics can be summarized by the following two equations:

$$y_k = m_k + \bar{\phi}_k \; D_k + \bar{\lambda}_k \; DV_k + \varepsilon_{y,k}, \tag{1}$$

$$m_k = m_{k-1} + \mu_k \; \tau_k + \phi_k (D_k - E_{k-1}[D_k]) + \lambda_k (DV_k - E_{k-1}[DV_k]) + \varepsilon_{m,k}, \tag{2}$$

where $\varepsilon_{y,k}$ and $\varepsilon_{m,k}$ denote the error terms, which we describe in detail below. All of the parameters are indexed by k, as they are updated with each new observation. In the transaction price dynamics (Equation (1)), $\bar{\phi}$ measures the transitory effect of the trade direction, while $\bar{\lambda}$ measures the transitory effect of the net order imbalance. In the fundamental value dynamics (Equation (2)), the first parameter μ_k measures the trade duration impact. The parameter ϕ measures the permanent effect of the trade direction innovation, while λ measures the permanent effect of the net order imbalance innovation. This formulation allows us to distinguish between the fixed effect of the order flow per trade (measured by ϕ and $\bar{\phi}$) and the variable effect per share traded (measured by λ and $\bar{\lambda}$) (see Glosten and Harris, 1988, and Sadka, 2006). The ultimate price impact of the trade will be a combination of these various effects.

Let us now describe how we model the error terms in Equations (1) and (2) to incorporate jumps. A key feature of our model is that jumps are allowed to affect both the transaction price and the fundamental value equations. Error terms are defined as:

$$\varepsilon_{y,k} = \sigma_{y,k}\ z_{y,k} + J_{y,k}, \tag{3}$$

$$\varepsilon_{m,k} = \sigma_{m,k}\ z_{m,k} + J_{m,k}, \tag{4}$$

where $z_{y,k}$ and $z_{m,k}$ denote the innovation terms, with $V[z_{y,k}] = V[z_{m,k}] = 1$, and $\sigma_{y,k}$ and $\sigma_{m,k}$ denote the time-varying variances of the continuous shocks, and $J_{y,k}$ and $J_{m,k}$ denote the transitory and permanent jumps, respectively. Permanent jumps will typically reflect firm or macroeconomic news, which often entail large price changes at the time they are released, whereas transitory jumps may be associated with temporary liquidity shortages.[1] Although price jumps are an important research topic in continuous time finance, they have been mostly ignored in the microstructure literature, perhaps due to the difficulty of filtering them out (see Duan and Fulop, 2007, for an exception). In our model, jumps may convey new information about the fundamental value of a stock. Moreover, these jumps are likely to contaminate the estimation of the model parameters and therefore must be accounted for.

Finally, both innovation terms $z_{y,k}$ and $z_{m,k}$ would ideally be Gaussian white noise. However, that is not likely to be the case for several reasons. First, Aït-Sahalia and Jacod (2009) have shown, from a theoretical perspective, that innovations could be generated by jumps with infinite activity, i.e., small but frequent jumps. As we are interested in filtering out jumps that may affect the estimation of the model parameters, we primarily focus on large but infrequent jumps. Second, empirically, price changes display some features that cannot be captured by a Gaussian distribution. For instance, tick-size discretization implies that some price changes are more likely than others. There is also a high frequency of zero price changes (approximately 50% in our data), which implies a peak at 0 in the distribution of price changes. We therefore expect the distribution of the transaction-price innovation to have fat tails. We discuss this point in more detail in Appendix 6.1.

[1] In this context, a liquidity shortage could, in principle, last for several trades before liquidity is restored. In this paper, we consider short-lasting shortages.

As the fundamental value is essentially driven by surprises in order flow, we now describe how we construct these innovations. To estimate the trade indicator conditional expectation, we rely on predictive regressions, as in Hasbrouck (1991), Foster and Viswanathan (1993), Brennan and Subrahmanyam (1996), and Sadka (2006). As a preliminary step, we consider a specification with past signed-volumes, past price changes, past trade-directions, and past durations:

$$\begin{aligned} DV_k &= E_{k-1}[DV_k] + \varepsilon_{V,k} \\ &= a_0 + \sum_{j=1}^{J} a_{1j} DV_{k-j} + \sum_{j=1}^{J} a_{2j} \Delta y_{k-j} + \sum_{j=1}^{J} a_{3j} D_{k-j} + \sum_{j=0}^{J} a_{4j} \tau_{k-j} + \varepsilon_{V,k}, \end{aligned} \quad (5)$$

where the residual term $\varepsilon_{V,k}$ represents the surprise in the signed trade volume.

The conditional expectation of the order flow is deduced from the trade volume Equation (5) by following the approach of Sadka (2006). The probability that the next trade will be buy-initiated, given the expected signed volume $E_{k-1}[DV_k]$, is $\Pr[D_k = +1 \mid E_{k-1}[DV_k]] = \Pr[E_{k-1}[DV_k] > -\varepsilon_{V,k}]$. If we denote by σ_V^2 the variance of $\varepsilon_{V,k}$, we find that the conditional probability of a buy order is $1 - F[-E_{k-1}[DV_k]/\sigma_V^2]$, where F is the cdf of the surprise in the signed volume. Assuming that this distribution is symmetric, $E_{k-1}[D_k] = 1 - 2F[-E_{k-1}[DV_k]/\sigma_V^2]$. We eventually compute the trade surprise as $\varepsilon_{D,k} = D_k - E_{k-1}[D_k]$.

Equations (1) and (2) represent a symmetric version of the model we investigate in this paper. In our general model, we separately consider the effect of buy-initiated and sell-initiated trades on both fundamental value and transaction price. Our final specification for the price dynamics can be summarized as follows:

$$y_k = m_k + \bar{\phi}_k^+ \; D_k^+ + \bar{\phi}_k^- \; D_k^- + \bar{\lambda}_k^+ \; DV_k^+ + \bar{\lambda}_k^- \; DV_k^- + \sigma_{y,k} \; z_{y,k} + J_{y,k}, \quad (6)$$

$$m_k = m_{k-1} + \mu_k \; \tau_k + \phi_k^+ \; \varepsilon_{D,k}^+ + \phi_k^- \; \varepsilon_{D,k}^- + \lambda_k^+ \; \varepsilon_{V,k}^+ + \lambda_k^- \; \varepsilon_{V,k}^- + \sigma_{m,k} \; z_{m,k} + J_{m,k}. \quad (7)$$

2.2 Estimation Methodology

The microstructure model described above has two main features, from an estimation perspective: first, it explicitly incorporates permanent and transitory jumps; second, the model parameters and innovation variances are time-varying. Our real-time estimation approach relies on particle filtering within a Bayesian framework, where we update the parameters not only on a daily basis but as observations materialize.[2] The estimation is based on the following steps:

1. Conditional on the information available at trade $k-1$ (parameters, variances, and state variables), we use the bootstrap filter of Gordon, Salmond, and Smith (1993) and Maiz, Miguez, and Djuric (2009) to determine the predictive density of the next transaction price y_k. This density is used to detect a jump at trade k.

2. If trade k is a jump, we do not update the model parameters and state variables. We then move to the next observation.

3. If trade k is not a jump, we update our estimates: Model parameters (μ, ϕ, λ, $\bar{\phi}$, and $\bar{\lambda}$) are updated in real time, using Bayesian OLS (BOLS), as soon as a new observation is available. The state variable (m_k) and variances ($\sigma^2_{m,k}$ and $\sigma^2_{y,k}$) are estimated via particle learning (Carvalho et al., 2010).

Jumps are detected in a non-parametric way, using an outlier detection procedure. The main motivation for such an approach is that it provides a simple way to differentiate between permanent and transitory jumps, as discussed in Fox (1972). A permanent jump is associated with a change in regime, whereas a transitory jump is a jump where only one trade's price deviates from the general trajectory. This distinction suggests a strategy for jump detection: Suppose at trade k, price y_k is far from the range of values implied by its distribution predicted using information available at trade $k-1$. If at trade $k+1$, the price y_{k+1} is back to a similar range of values from where it started, we define it as a transitory jump. If y_{k+1} remains close to its previous position, we define the jump at k as a permanent jump. With this strategy, jumps are detected in real time with a maximum of a one-trade lag, so that the classification of the jumps will require only one additional observation.

[2]Recent surveys of these techniques are found in Doucet and Johansen (2009) and Lopes and Tsay (2011).

To decide if the price at trade k is a jump, we construct a confidence interval based on the predictive distribution of price at trade k, conditional on the information up to trade $k-1$. Given the non-Gaussianity of the innovation terms even after removing large jumps, the predictive distribution is constructed to reflect the actual properties of the price process distribution. Our approach is to assume a Student t distribution, whose degree of freedom is equal to the inverse of the tail index estimated over the data. Once the predictive distribution for trade k is estimated, we define the thresholds for jump detection as the $(\alpha/2)$ quantile for the negative jumps and the $1-(\alpha/2)$ quantile for the positive jumps.[3] See Appendix 6.1 for more details on the estimation methodology.

2.3 Simulations

As detailed in the algorithm above, jumps are detected in a non-parametric way, using an outlier detection procedure. In this short section, we illustrate how this strategy works on simulated data. We consider the following simplified data generating process:

$$y_k = m_k + \sigma_y \; z_{y,k} + J_{y,k},$$
$$m_k = m_{k-1} + \sigma_m \; z_{m,k} + J_{m,k},$$

where innovations $z_{y,k}$ and $z_{m,k}$ are uncorrelated i.i.d. $N(0,1)$ processes and jumps $J_{y,k}$ and $J_{m,k}$ are independent compound Poisson processes. When there is no jump, $J_{y,k}=0$ and $J_{m,k}=0$. When there is a jump, it will be drawn from a Gaussian distribution. The intensity of the jumps is defined as follows: $J_{y,k}$ takes a non-zero value with an intensity of $\lambda_y = 1/50$ and $J_{m,k}$ has an intensity of $\lambda_m = 1/50$. If a jump occurs, then $J_{y,k} \sim N(2,1)$ and $J_{m,k} \sim N(-2,1)$.

We simulate a sample with $T = 200$ observations starting with $m_0 = 100$ and we set $\sigma_y = 20\%$ and $\sigma_m = 10\%$.[4] Such a magnitude for the standard deviations is consistent

[3] In the empirical application, we consider $\alpha = 0.2\%$, meaning that we classify as jumps all the realizations below the 0.1% quantile or above the 99.9% quantile.

[4] This illustration focuses on describing outlier treatment and abstracts intraday volatility variation for exposition parsimony. We note that in general, outlier detection is contingent upon local volatility estimation. In our empirical application, we allow volatility to vary intradaily. An exhaustive Monte-

with our actual data.[5] Eventually, we focus on a window covering observations 100 to 200 and apply the algorithm for the jump detection and the parameter estimation described in Section 2.2 and Appendix 6.2. The results are presented in **Figures 1 and 2**.

Let us start with the discussion of Figure 1. The observations are represented by an o symbol and the true values of the state are represented by a +. At each step, the particle filter provides us with the median estimate of the state. This is represented by the continuous line. We notice that this line tracks very well the actual states. We also represent a confidence interval following our modification of the jump detection algorithm of Maiz, Miguez, and Djuric (2009). This confidence interval is represented with dashed thin lines.

At observation 133, there is a large negative jump. The algorithm detects this jump and indicates that it is a permanent (or innovation) jump. This is followed by an immediate adjustment, backwards, of the estimation of the state once a new observation becomes known. Because of this backwards step, the estimation of the state is adjusted as can be seen by inspecting the continuous line, which touches the center of the circle (the cross would not be known in a real life exercise since this is the latent state).

At observations 101, 124, and 159, we have large positive observations. In these cases, the jump detection algorithm calls for transitory (or additive) jumps. The observations are ignored and the state is not updated, which translates into a small horizontal step in terms of the underlying state estimation and its associated confidence intervals.

Figure 2 corroborates those findings. The upper figure presents the distance between the actual observations to the filtered estimates. For observation 133 where a jump was detected, given the way that the algorithm performs the correction of the state estimation, we find no change in the observation equation. In observations 101, 124, and 159, we have additive outliers, which the algorithm neglects in the estimation of the state. As the state is not updated, we obtain large differences between the observation y_k and the retained state m_k.

Carlo exercise, including various data generating processes and estimation methods, is of interest and left for future research.

[5] We experienced with various signal-to-noise ratios and various parameters and found our method to perform well in all the experiments we ran.

If we turn to the lower figure, we verify that, for observation 133, there is a relatively large change in the state ($m_k - m_{k-1}$), which results from the correction that was made in recognition of the jump. On the other hand, for innovations 101, 124, and 159, there is no variation in the state, as the algorithm recognized that there was a transitory jump in the observation and therefore decided not to update the state.

3 Data and Preliminary Analysis

3.1 Data

Euronext-Paris is an electronic limit order market based on a trading platform called "Nouveau Système de Cotation (NSC)." Limit orders consist of a limit price and a quantity to buy or sell at that price. Orders are submitted by investors through brokers and stored in the limit order book. The matching of orders and corresponding trades follows strict price and time priorities. Market orders (i.e., orders to buy or sell at the best price available in the book) are immediately executed and matched with the best orders on the opposite side of the book. The exchange opens at 9:00 am and closes at 5:30 pm. In this market, liquidity is provided by the limit order book, as there are no market makers. The functioning of this system has been described in detail in Biais, Hillion, and Spatt (1995), De Jong, Nijman, and Röell (1996), and Foucault, Moinas, and Theissen (2007).

The data consist of all transaction prices and quantities from the Euronext-Paris "BDM" database. The sample used consists of 12 constituent stocks of the CAC40 index plus one stock that had just exited the index, for the months of January and February of 2003, i.e., 42 trading days.[6] We have chosen stocks to represent various industries and a wide selection of characteristics, enabling us to investigate the relationship between price impact and common proxies such as number of trades and average durations between trades. Transaction prices were adjusted in the following way: trades before 9 am and after

[6]We also investigated more recent data. One difficulty we encountered was the explosion in the number of trades from 2005 onward. For the highly liquid stock, Total, the number of trades per day has risen from 3,440 in 2003 to 9,750 in 2006 and 18,700 in 2009. With so many trades, our procedure still works but is more time consuming. Additionally, owing to algorithmic trading, the dynamics of the price of highly liquid stocks has changed. We view our model as applying to stocks with an average liquidity. (Highly illiquid stocks are auctioned at Euronext-Paris.)

5:30 pm were eliminated. The opening trade was also eliminated because it corresponds to an auction. Trades occurring at the same second but at different prices are aggregated with a single price given by the weighted average of the prices.

We use the best bid and best ask quotes to determine trade direction. If the transaction price is at or above the best ask quote, the trade is classified as a buy-initiated trade. Symmetrically, if the price is at or below the best bid quote, the trade is classified as a sell-initiated trade. In our dataset, only a small percentage of trades (approximately 4%) was not classified. Venkataraman (2001) also uses data similar to ours and reports that data from the Paris Bourse are relatively error free as they are produced by the automated trading system. For unclassified trades, we applied the Lee and Ready (1991) classification algorithm.[7] Using the bid and ask before a trade, we compute the midquote. If the trade price is above (below) the midquote, the trade is classified as a buy (sell) order. If the trade price is at the midquote but higher (lower) than the previous traded price, it is classified as a buy (sell) order. If the trade price is at the midquote and equal to the previous traded price, then we go further back in time and select the first transaction price which differs. If the current price is higher (lower) than this price, then the trade is a buyer (seller) initiated order.

Table 1 provides some summary statistics of the 12 stocks during the sample period. The total number of trades varies greatly across stocks, revealing the wide spectrum of liquidity we consider in our application. Liquidity can also be measured by the average duration between trades. Average duration ranges from 6.7 seconds between two trades for Alcatel to more than 40 seconds for Sodexho (median durations are 4 seconds and 15 seconds, respectively). The average monetary trading volume ranges between 7,500 euros for Alstom and 110,460 euros for Total (median values are 2,700 euros and 54,000 euros, respectively). Based on market capitalization and number of trades, we will classify Total, France Telecom, and AXA as "highly liquid" stocks, Alstom, Lagardere, and Sodexho as "less liquid" stocks, and the remaining stocks as "intermediate".[8]

[7] In an empirical analysis involving Nasdaq data, Ellis, Michaely, and O'Hara (2000) report correct classification with this algorithm of 80% of the trades.

[8] Alstom left the CAC40 several months before the period under consideration. Orange was a subsidiary of France Telecom and had a relatively narrow free float, explaining its low number of trades despite its large market capitalization. Alcatel is in the opposite situation, with a low market capitalization but the

We also observe wide diversity with respect to the properties of price changes. **Figure 3** displays the evolution, over five days in January 2003, of the price and price increments of two stocks. Starting with top figures, which display the price process, we observe very large price variations, particularly for the less liquid stock, Sodhexo. For the liquid stock Total, the largest price changes are −46 bp and +40 bp (approximately to 0.3% of the price) (Figure 3-a). For Sodexho, if we exclude opening price changes, the largest price changes are −48 bp and +41 bp (approximately 2% of the price) (Figure 3-b). As bottom figures reveal, the properties of the price increments differ significantly for highly liquid and less liquid stocks. For Total, the price discreteness is evident (Figure 3-c), whereas for Sodexho, whose price dynamics is mostly dominated by the duration between trades, price discreteness is much less evident (Figure 3-d).

The standard deviations of raw returns range from 6.2 bp for Total to more than 25 bp for Sodexho. If returns are expressed in business time, standard deviations decrease significantly to 3.5 bp and 10.7 bp, respectively. Comparison between unscaled and scaled returns demonstrates the importance of the duration between trades in price formation. A large portion of the price volatility (here between 2 and 3 times) is driven by discontinuities in the price process. These findings demonstrate, in our view, the importance of taking into account durations between trades in price formation.

As the skewness and kurtosis of the intraday returns reveal, stock returns are highly non-normal. For instance, Suez exhibits a highly rightward skewed distribution, whereas Vivendi exhibits a highly left-skewed distribution. All stock distributions are characterized by fat tails, as confirmed by large kurtosis values. The tail index also ranges from 0.24 to 0.47, indicating that, for some stocks, the kurtosis and (perhaps) the skewness are infinite.[9]

largest number of trades during the sample period. One explanation is the low price of a share, enabling very small trades to take place. This is confirmed by the low trade size of Alcatel (second lowest median monetary trade size).

[9]Our estimates are based on the average of the left and right Hill's estimators of the tail index with a threshold of 5% of the observations. A tail index of ξ indicates that the moments above $1/\xi$ may not be defined.

3.2 Intradaily Seasonality

Intradaily seasonality patterns of market variables have been highlighted in several papers, including Andersen and Bollerslev (1997 and 1998) for return volatility, Jain and Joh (1988) and Foster and Viswanathan (1993) for trading volume, as well as Engle and Russell (1998) and Dufour and Engle (2000) for duration between trades. It is therefore important to remove this systematic component before estimating the microstructure model to avoid biases in the estimation of the model parameters.[10]

We use a specification similar to Engle and Sun (2007) to obtain deseasonalized price increments. The volatility of the fundamental value is defined as $\sigma^2_{m,k} = \delta^2_i \; \tilde{\sigma}^2_{m,k}$, where δ_i captures the seasonality effect and $\tilde{\sigma}^2_{m,k}$ is the variance of the kth return conditional on information available up to trade $(k-1)$. Appendix 6.2 describes the robust method we use to estimate δ_i, along the lines of Boudt, Croux, and Laurent (2011).

We average the seasonality coefficients δ_i across all firms to capture the main trend, and smooth the resulting profiles using the Loess algorithm. Price increments are standardized by their corresponding averaged and smoothed seasonal component δ_i. These periodicity-free returns are then used to estimate the dynamic path of $\tilde{\sigma}_{m,k}$ using the algorithm described in Appendix 6.1.

For microstructure variables, seasonality is obtained as follows. We consider $M = 51$ trading intervals i of 10 min (from 9 am to 5:30 pm), $i = 1, \cdots, M$ and compute time-of-the-day seasonality coefficients for each interval i. For the duration between trades, the log trading volume (number of shares), and the log monetary volume, we estimate the seasonality coefficients as the average value of the variable over the trading interval i. As for periodic volatility, we use the average across firms, smoothed with the Loess algorithm. Once we have obtained seasonal coefficients, we rescale each new microstructure observation.

[10] MRR (1997) adopt a different strategy to address seasonality. They estimate their model for different intraday trading intervals. They also find significant differences in microstructure parameter estimates across the intervals.

To maintain real-time estimation, we only use the first 5 days of the sample to estimate the seasonality coefficients.[11]

In Figure 4, we display the intraday seasonality of price change volatility, the duration between trades, the log trading volume (number of shares), and the log monetary volume. We observe a U-shaped pattern for return volatility (a pattern found in several previous papers) and an inverse-U-shaped pattern for duration between trades. We also find that trading volume increases almost continuously during the day.

3.3 Order Flow Surprises

We considered several specifications to predict the order flow. We eventually retained a model with 10 lags of signed trade volume, price changes, trade direction, and duration between trades.

Table 2 presents the regression results. To conserve space, we report the sum of the 10 parameters and the standard error corresponding to the sum of the parameters. Signed trade volume is somewhat persistent. The sum of the coefficients of the first 10 lags ranges between about 0.1 and 0.4 across firms and is always highly significant. The additional contribution of trade direction (D_{k-j}), conditional on past volumes, is unclear. The sum of the parameters is in general statistically significant, although signs vary among stocks. As suggested by Foster and Viswanathan (1993), past returns contribute to the prediction of subsequent volumes, with a coefficient that is positive for all stocks except one and significant for six stocks. Finally, the duration between trades has a significant and positive impact on signed volume, suggesting that the larger is the time interval since the last trade, the higher is the expected volume. The R^2 of the regressions ranges from 1%, for highly liquid stocks (such as Total or France Telecom), to above 5%, for less liquid stocks (such as Lagardere or Sodexho).

[11] We experimented several alternative approaches: we used a weighted and unweighted average of seasonality coefficients across firms; we used alternative smoothing algorithms, such as Hodrick and Perscott or Golay and Solvay; we used different values for the smoothing parameter in Loess algorithm; we used different subsamples (5 days, 10 days, all days) to estimate the seasonality coefficients. We found no material differences in the subsequent estimates.

3.4 Model with Constant Parameters and no Jump Filtering

We start with the model with constant parameters and variances and no jump filtering. The assumptions of the model are close to those of Sadka (2006). **Table 3** reports the parameter estimates for the model both with (Panel A) and without the microstructure variables (Panel B).[12] Comparison between the two models clearly indicates that microstructure variables help explaining the dynamics of the stock return. On average, the variance of the transaction price innovation (σ_y^2) is reduced by 20%. The likelihood-ratio test statistics for the null hypothesis that the microstructure variables have no effect (LR_2), reported in the last row of the table, are all very large and significant. They indicate that the microstructure variables should be included in the model.

The relevance of the order flow is also revealed through the significance levels and magnitudes of the parameters in the microstructure model. We find that, in the fundamental value equation, parameters associated with the permanent effect of order flow (ϕ^+, ϕ^-, λ^+, and λ^-) are all highly statistically significant and have the expected signs: ϕ^+ and ϕ^- lie between 1.3 and 6.5, and λ^+ and λ^- lie between 0.3 and 2.3. The large positive values of parameters ϕ^+ and ϕ^- indicate that surprises in trade direction have permanent effects on prices, irrespective of the magnitudes of the trades. This effect is more pronounced for less liquid stocks. Similarly, the positive signs of λ^+ and λ^- indicate that an unexpectedly large trading volume also has a large price impact.

Turning to the transaction price equation, we observe significant deviations from the equilibrium price due to trading activity. Parameters $\bar{\phi}^+$ and $\bar{\phi}^-$ range from 1.5 to 11.2, whereas parameters $\bar{\lambda}^+$ and $\bar{\lambda}^-$ range from -2 to -0.3. The positive sign of parameter $\bar{\phi}$ suggests that the order flow has a large transitory price impact that eventually reverts to its long-run level given by ϕ. On the other hand, the universally negative parameter $\bar{\lambda}$ demonstrates that the transitory price impact of large volumes is negative. This result suggests a compensation effect between the fundamental value and transaction price dynamics. The price is not instantaneously affected by the magnitude of a sell-initiated trade, as the cumulative effect of the transitory and permanent impacts is almost 0. However,

[12] The model is estimated with the standard Kalman filter. All of the parameters reported in the table are multiplied by 100 for ease of reading and can be directly interpreted in basis points.

the price will eventually converge to the new fundamental value, which incorporates the negative impact of the sell trade. The same reasoning holds for buy-initiated trades. These parameter estimates accord with the estimates of Sadka (2006) for NYSE stocks.

The table also reports the likelihood-ratio test statistics for the null hypothesis that buy- and sell-initiated trades have the same price impact (LR_1). The 1% critical level is exceeded in all but three cases. Thus, buy orders and sell orders will generally differ in their price impacts. We investigate this asymmetry pattern in more detail in Section 4.4.

4 Microstructure Model and Price Impact of Trades

We now turn to the estimation of the complete microstructure model with time-varying parameters and jump filtering. We start by presenting some results about the jumps detected with our particle filter approach. We then present the results for the complete model with time-varying parameters and jump filtering. This allows us to discuss the dynamic patterns of the price impact of trades.

4.1 Jump Analysis

As we explained above, we have a jump at trade k if the price change is below the 0.1% or above the 99.9% quantiles of the predictive distribution (i.e., conditional on the information available at trade $k-1$). As detection is performed in real time, it is possible that systematic over-detections or under-detections occur. Another important feature of our approach is that the magnitudes of the jumps depend on the properties of the price changes and may therefore differ markedly from one stock to another.

Summary statistics of the jumps detected by our estimation procedure are reported in **Table 4**. A first general comment is that there are wide disparities in the numbers of jumps across stocks: The average number of jumps per day ($J_y^+ + J_y^- + J_m^+ + J_m^-$) varies from 22, for Alcatel, to just one, for Sodexho (Panels A and B). It is worth emphasizing that this difference is due not only to the difference in the total number of trades (Alcatel has only 6.5 times more trades than Sodexho). Indeed, it turns out that the frequency of jumps per day differs significantly between highly liquid stocks and less liquid stocks (Panel

C). For highly liquid stocks, we find a frequency of jumps that is slightly above the ex-ante threshold we use in the detection algorithm ($\alpha = 0.2\%$): we find a frequency of 0.6% for Orange, 0.5% for Alcatel and AXA, and 0.4% for France Telecom, Total, and Vivendi. For less liquid stocks, the algorithm detects the expected number of jumps: approximately 0.1%–0.2% for Alstom, Lagardere, LVMH, and Sodexho.[13]

Another important result concerns the magnitude of jumps (Panel D). If we consider highly liquid stocks, we observe that the average jump size is relatively small. It is below 20 bp for Total, on average, although only 0.38% of trades in this stock are considered jumps. For most of the other highly liquid stocks, the average jump size is 30–40 bp.[14] For less liquid stocks, jumps are much larger, approximately 90 bp for Alstom and Lagardere and 110 bp for Sodexho.

There are also some similarities and differences between transitory and permanent jumps. On the one hand, on average, the proportion of jumps per day is the same for the two types of jumps. There are 30% more transitory jumps for Alcatel and Orange, but 40% less for STMicro. On the other hand, transitory jumps are on average larger than permanent jumps. This is particularly the case for Alstom and Total (approximately 75% more transitory jumps).

We also investigated the intradaily seasonality of detected jumps by counting, for all companies, the numbers of certain types of jumps that occurred during various hours of the day. The results of this investigation are presented in **Table 5**. The lower panel of the table is constructed as follows: we denote by N_{idh} the number of jumps found for company i, on day d and in hour h. We then define by $T_h = \sum_i \sum_d N_{idh}$ the total number of jumps for a given hour h over the sample. Eventually, we construct $f_h = 100 \times \frac{1}{12} \sum_{i=1}^{12} \sum_d N_{idh}/T_h$, which is the average number of jumps across all companies for a given hour h. As can be

[13] A possible interpretation of this difference is as follows: The distribution used to define the quantile is based on observed price changes. The algorithm then defines the quantiles of the distribution of the innovation ($\varepsilon_{y,k}$), i.e., after incorporating the effects of the microstructure variables and the time-varying volatility. As the model with explanatory variables performs relatively well for highly liquid stocks, the tails of the innovation distribution are thinner than those of the observed price changes, and therefore more observations are classified as jumps.

[14] It should be noted that, for some stocks, the tick size represents a significant fraction of the price level. For Alstom, Alcatel, and Orange, a variation of 1 cent would correspond to a price change of 23 bp, 16 bp, and 13 bp, respectively. Conversely, for these three firms, an average jump (1%, 0.4%, and 0.5%, respectively) would correspond to a price change of 4.4 cents, 2.5 cents, and 9 cents, respectively.

seen, the number of jumps is especially high during the opening and the closing hours.[15] As it is well documented, these are the periods when trading activity is most intense, as also confirmed by the seasonality pattern for the price change volatility (Figure 4). Our detection of jumps during these moments suggests that more news is generated at these times. Thirty percent of the detected jumps occur during the first hour of trading, whereas only 4.3% of the jumps are found to occur between 1 pm and 2 pm (see Table 5, last column). In the afternoon, the frequency of jumps increases again, and 15% of the jumps occur between 4 pm and 5 pm. On average, jumps are 7 times more likely during the first hour than during the fifth hour.

The breakdown into transitory jumps J_y and permanent jumps J_m also provides interesting results. If jumps in the state equation can be associated with fundamental news, our estimations show that news early in the morning and late in the afternoon is particularly relevant for the Paris Bourse. Fifty percent of jumps occurring in the first hour are classified as permanent. Between 1 pm and 2 pm, only 42% of jumps are considered permanent. Finally, the proportion of jumps increases again at the end of the day, to 45%, between 4 pm and 5 pm.

4.2 Model with Time-Varying Parameters and Jump Filtering

We now consider the model with dynamic parameters and variances and preliminary jump filtering. **Table 6** presents averages of the parameters over the sample and the standard deviations of the parameter series.[16] Comparison of the models with and without microstructure variables (Panels A and B, respectively) shows that these variables significantly contribute to the dynamics of stock returns. On average, the variance of the transaction price innovation (σ_y^2) is reduced by 22%. Overall, microstructure variables

[15]The last subperiod lasts only 30 minutes (from 17 to 17:30), so that the numbers are not directly comparable.

[16]The model is estimated using the particle filter. Again, all the parameters are multiplied by 100. Clearly, the standard deviations associated to the parameters are not comparable to the standard errors reported for the Kalman filter estimation. Instead, they reflect the magnitudes of the changes in the dynamics of the particle filter parameters. Large standard deviations indicate that allowing for time variability in the model parameters is, in fact, important in obtaining a good description of actual data.

(with time-varying parameters) and jump filtering reduce the variance of the transaction price innovation by 30%.

Comparison of these averages with the estimates of Table 3 reveals large differences between the two models, implying that dynamic parameters and jump filtering significantly affect the estimation of the microstructure model and thus the measure of the price impact of order flow. In the fundamental value equation, the parameters ϕ^+ and ϕ^-, which reflect the direct permanent price impact of trades, are significantly higher. The increase in parameters ranges from 0.4 to 1.7, suggesting that after removing jumps, the price dynamics is more sensitive to trades. This effect is sizable, as the permanent price impact increases from 6% to 101%. We will show below that the magnitude of the effect has important economic implications. Removing jumps also implies, for most stocks, an increase in the transitory price impact (parameters $\bar{\phi}$ and $\bar{\lambda}$), although to a lesser extent. Finally, the impact of the duration between trades is considerably reduced and is indeed halved for most stocks. In sum, the comparison suggests that filtering out jumps enables us to better capture the microstructure phenomena involved in price formation.

Figures 5 and 6 present the evolution of the parameters of a highly liquid stock (Total) and a less liquid one (Sodexho). The constant lines are the Kalman filter estimates. All of the parameters are multiplied by 100 and thus correspond to price impacts measured in basis points. Starting with Total (Figure 5), we observe that the parameters are relatively stable, varying within a relatively narrow band. For instance, the four ϕ parameters range from 1 bp to 3.5 bp. On the opposite side, Figure 6 shows that, for the less liquid stock, there is much greater time variability in the model parameters. The ϕ parameters vary between -5 bp and 15 bp.

Inspecting the figures corresponding to buy and sell trades, we observe some divergence between the parameters. Inspection of the average of the parameter estimates suggests that the asymmetry in the price impact between buy- and sell-initiated trades comes mainly from the differences in the permanent effect of the order flow: we find that, in general, $\phi^+ > \phi^-$ for highly liquid stocks and $\phi^- > \phi^+$ for less liquid stocks. We also observe that, in general, $\bar{\lambda}^+ > \bar{\lambda}^-$ for highly liquid stocks and $\bar{\lambda}^- > \bar{\lambda}^+$ for less liquid stocks, suggesting differences in the transitory effects of trading volume. This difference translates into differential price impacts of trade, as we will now see.

4.3 Contribution Analysis

With our complete model, we can analyze the contribution of the various factors driving the dynamics of the transaction price. To simplify notations, we rewrite our final specification for the price dynamics (Equations (6) and (7)) as follows:

$$y_k = m_k + Z_{1,k}\ \beta_{y,k} + \eta_{y,k} + J_{y,k}, \tag{8}$$

$$m_k = m_{k-1} + Z_{2,k}\ \beta_{m,k} + \eta_{m,k} + J_{m,k}, \tag{9}$$

where $Z_{1,k} = [\ D_k^+ \quad D_k^- \quad DV_k^+ \quad DV_k^-\ ]$ and $Z_{2,k} = [\ \tau_k \quad \varepsilon_{D,k}^+ \quad \varepsilon_{D,k}^- \quad \varepsilon_{V,k}^+ \quad \varepsilon_{V,k}^-\ ]$, $\eta_{y,k} = \sigma_{y,k}\ z_{y,k}$, and $\eta_{m,k} = \sigma_{m,k}\ z_{m,k}$. The dynamics of the transaction price can be written as:

$$\begin{aligned} y_k - y_{k-1} &= (m_k - m_{k-1}) + (Z_{1,k}\beta_{y,k} - Z_{1,k-1}\beta_{y,k-1}) + (\eta_{y,k} - \eta_{y,k-1}) + (J_{y,k} - J_{y,k-1}) \\ &= Z_{3,k}\ \beta_{ym,k} + (\eta_{y,k} - \eta_{y,k-1}) + \eta_{m,k} + (J_{y,k} - J_{y,k-1}) + J_{m,k}, \end{aligned}$$

where $Z_{3,k} = [\ Z_{1,k} \quad -Z_{1,k-1} \quad Z_{2,k}\ ]$ and $\beta_{ym,k} = [\beta_{y,k} \quad \beta_{y,k-1} \quad \beta_{m,k}]'$ denote the vectors of explanatory variables and structural parameters, respectively. We notice that, by construction of our algorithm, jumps of different kinds are not correlated and jumps are not correlated with continuous innovations $\eta_{y,k}$ and $\eta_{m,k}$. In contrast, $\eta_{y,k}$ and $\eta_{m,k}$ can be correlated.

Denoting $r_k = y_k - y_{k-1}$ the return of the stock at trade k, the total variation in transaction return is given by:

$$\begin{aligned} V[r_k] &= V[Z_{3,k}\ \beta_{ym,k}] + V[\eta_{y,k} - \eta_{y,k-1}] + V[\eta_{m,k}] + 2Cov[\eta_{y,k} - \eta_{y,k-1}, \eta_{m,k}] \\ &\quad + V[J_{y,k} - J_{y,k-1}] + V[J_{m,k}]. \end{aligned}$$

The first term on the right-hand side corresponds to the contribution of the microstructure model, the second to fourth terms correspond to the contribution of the continuous innovations, and the fifth and sixth terms correspond to the contribution of the transitory and permanent jumps, respectively. To be comparable with the usual definition of the return process, we need to take the contribution of the overnight return into account. Therefore, we consider $V[r_k]$ as the variance over all the trades, including the overnight return.

As **Table 7** shows, the contribution of the intraday jumps is relatively limited: it is 8.5% on average, with a maximum of 12.2% for Orange and a minimum of 5% for Alstom. This contribution is in fact similar to the contribution of the overnight return. On average, the overnight return contribute for 7.1% of the total variation, with a maximum of 15.4% for STMicro and a minimum of 3.7% for Alstom.

These results can be compared to previous estimates of the jump contribution reported in the literature. Estimates based on similar data (large-capitalization stocks) suggest a large frequency effect: using daily data, Ball and Torous (1985) find a contribution of jumps of 47% on average (sample: 1981–1982) and Maheu and McCurdy (2004) find an average contribution of 29% (sample: 1962–2001). Using intraday (17.5 min) returns, Bollerslev, Law, and Tauchen (2008) find a contribution of 12% (sample: 2001–2005). Christensen, Oomen, and Podolskij (2014) also compare, on the same data set, the jump contribution depending on the sample frequency. On DJIA constituents, they find a contribution of 11.9% for 15 min sampling, 7.3% for 5 min sampling, and finally 1.3% for tick frequency. All these estimates are based on non-parametric techniques. Although our results are based on a microstructure approach, our estimates are, therefore, in the ball park of those reported by Christensen, Oomen, and Podolskij (2014).

The relative contributions of the transaction price innovation and fundamental value innovation are rather different from one stock to the other. For instance, for STMicro, LVMH, and Sodexho, the contribution of the fundamental value innovation is close to 20% of the total return variation, a proportion that is larger or equal to the contribution of the transaction price innovation. For these firms, most of the continuous volatility comes from the volatility in the fundamental value. In contrast, for stocks such as Alcatel, Alstom, or Suez, the contribution of the continuous innovation is dominated by the transaction price innovation.

We now consider the contribution of the microstructure model. The average contribution is equal to 47.7%, with large contributions for Alcatel, Alstom, and Total (above 58%) and low contribution for STMicro, Sodexho, and Lagardere (below 40%). These results suggest that the microstructure model is able to capture a significant part of the variation in stock returns at tick frequency.

4.4 Price Impact of Trades

Several papers have investigated the price impact of block trades of institutional investors (Holthausen, Leftwich, and Mayers, 1987, Chan and Lakonishok, 1993, Keim and Madhavan, 1995), almost universally reporting that buy trades have larger price impacts than sell trades. The dominant explanation for this finding is based on the short-sale constraints faced by institutional investors. A decision to buy a stock implies that the institution has a strong positive opinion about the stock relative to the universe of available stocks. In contrast, a decision to sell a stock implies a negative opinion about the stock relative to the stocks held by the institution. Saar (2001) proposes a theoretical model in which this asymmetry can be reversed. The main implication of his model is that a long run-up in a stock's price reduces the asymmetry in the permanent price impact of buy-initiated and sell-initiated block trades.

We now investigate this question within our framework by considering all the trades of a stock instead of the block trades only. For this purpose, we consider large buy-initiated and sell-initiated trades, equal to 10 times the average trade of the stock (from $V_k = 75,000$ euros for Alstom to $1,105,000$ euros for Total). We then evaluate the expected (permanent and transitory) impact on price. For buy- and sell-initiated trades, we have the following equations for the permanent (PPI) and transitory (TPI) price impacts, respectively:[17]

$$PPI^+ = \phi^+ + \lambda^+ \, \varepsilon^+_{V,k}, \qquad PPI^- = -\phi^- + \lambda^- \, \varepsilon^-_{V,k}, \tag{10}$$

$$TPI^+ = \bar{\phi}^+ + \bar{\lambda}^+ \, DV^+_k, \qquad TPI^- = -\bar{\phi}^- + \bar{\lambda}^- \, DV^-_k. \tag{11}$$

[17] We assume that the expected trade direction is 0 and that the expected trade size is the average trade size for the given stock over the sample.

This decomposition is similar to that found in Sadka (2006).

Table 8 reports several important results. First, the price impact estimates are dramatically altered when we move from a model with constant parameters and no jump filtering (Kalman filter) to a model with time-varying parameters and jump filtering (particle filter). For most of the firms, the price impact is increased by approximately 10 to 30% compared to Kalman filter estimates. Failing to take parameter dynamics and jumps into account therefore entails systematic underestimation of the price impact.

Second, focusing now on particle filter estimates, we observe that there are large differences in price impacts across firms. A generally expected result is that highly liquid firms are less affected by trades of a given value than less liquid firms. A 10 times larger than average trade would change the price of Alstom by as much as 13 bp if the trade is buy-initiated and 18 bp if it is sell-initiated. On the other side of the spectrum, the price impact of such a trade on Total is only approximately 4 bp regardless of the side of the trade.

Third, we find, as expected, an asymmetry in the price impact of buy-initiated trades and sell-initiated trades, but this asymmetry can also be reversed. Buys have a larger impact than sells for highly liquid stocks with a large proportion of sell-initiated orders. In contrast, sells have a larger impact than buys for less liquid stocks with a large proportion of buy-initiated orders. This result holds for the overall price impact but also for the permanent and transitory components. This finding is consistent with the information channel, described theoretically by Saar (2001), through which greater trading intensity and more informative trades are likely to increase the asymmetry of the price effects.

To illustrate this result, **Figure 7** displays the total price impact of a buy-initiated trade and a sell-initiated trade for two highly liquid stocks (Total and France Telecom) and two less liquid stocks (Alstom and Sodexho). The straight lines correspond to the Kalman filter case, indicating that price impact is almost systematically under-estimated under this specification. Buy trades have a greater impact than sell trades for highly liquid stocks and a smaller impact for less liquid stocks. It should be noticed that the reverse asymmetry does not hold for all less liquid stocks: The usual asymmetry holds for Alstom and Sodexho.

5 Conclusion

In this paper, we have considered a market microstructure model in which stock price is modeled as gravitating around the fundamental value process. Differences between transaction price and fundamental value are explained by trade direction and trade size, whereas changes in the fundamental value are explained by surprises in trade direction and trade size. As price changes are also affected by (permanent and transitory) jumps, we develop an estimation strategy based on Bayesian OLS and particle filtering that allows us to detect jumps in real time.

In the empirical section of this paper, we estimate the model, using all trades of 12 large stocks on Euronext Paris over a two-month period. The jumps detected by our estimation strategy have interesting properties. First, highly liquid stocks are characterized by several relatively small jumps per day, whereas less liquid stocks are characterized by a small number of relatively large jumps per day. Additionally, less liquid stocks tend to exhibit a larger proportion of permanent jumps than highly liquid stocks. Finally, jumps of both types tend to occur most frequently during the first and final hours of the trading day. This result suggests that jumps may partly trigger volatility, which displays a similar seasonality pattern. The impact of jumps at tick frequency is relatively limited, however, as their overall contribution to the total variation of stock returns is equal to 8.5% on average. This contribution is similar to the contribution of the overnight jump (7% on average). This result is consistent with the evidence reported by Christensen, Oomen, Podolskij (2014), using a non-parametric measure of jumps at tick frequency. The contribution of the continuous innovations is approximately 36.7% of the total variation of stock returns, whereas the microstructure model explains on average 47.7% of the total variation.

Furthermore, we obtain several important results regarding the price impact of trades. First, measures of the price impact are sensitive to the dynamics of the model's parameters and the preliminary detection of jumps. Indeed, explicitly modeling the microstructure variables significantly reduces the variance of the innovation processes. In addition, allowing for dynamic parameters and jumps strongly affects the estimated price impact of trades.

In our model with jump filtering, we find that the price impact is 10 to 30% higher than in our model without jump filtering. Second, as expected, less liquid stocks are more sensitive than highly liquid stocks are to trades. However, we also identify an asymmetry in the price impact of buy- and sell-initiated trades: highly liquid firms with a large proportion of sell-initiated orders are more sensitive to buy-initiated trades, whereas less liquid firms with a large proportion of buy-initiated orders are more sensitive to sell-initiated trades.

6 Appendices

6.1 Appendix 1: Estimation Methodology

To simplify notations, we rewrite our final specification for the price dynamics (Equations (6) and (7)) as follows:

$$
\begin{aligned}
y_k &= m_k + Z_{1,k}\ \beta_{y,k} + \sigma_{y,k}\ z_{y,k} + J_{y,k}, && (12)\\
m_k &= m_{k-1} + Z_{2,k}\ \beta_{m,k} + \sigma_{m,k}\ z_{m,k} + J_{m,k}, && (13)
\end{aligned}
$$

where $Z_{1,k} = [\ D_k^+ \quad D_k^- \quad DV_k^+ \quad DV_k^-\ ]$ and $Z_{2,k} = [\ \tau_k \quad \varepsilon_{D,k}^+ \quad \varepsilon_{D,k}^- \quad \varepsilon_{V,k}^+ \quad \varepsilon_{V,k}^-\ ]$. The time index on $\beta_{y,k}$ and $\beta_{m,k}$ reflects the fact that parameters will be updated with each new observation.

To start the discussion about the estimation of the model, we consider a simplified model without jumps. Once this simpler model has been discussed, we will turn to the detection and treatment of the jumps (Section 6.1.3). In order to take parameters variability into account, we re-initialize the parameters each day and use each new observation y_k to update the parameter estimates. Such an approach is referred to as online estimation in the particle filter literature.

6.1.1 Updating Parameters with Bayesian OLS

Before discussing the parameter-learning algorithm for the estimation of the state variable and the continuous shock variances, we describe the way we update the estimates of $\beta_{y,k}$ and $\beta_{m,k}$. We assume that N_d trades are available on a given day d. In a traditional OLS setting, we would simply estimate β_y and β_m from the regression:

$$
y_k - y_{k-1} = (Z_{1,k} - Z_{1,k-1})\beta_y + Z_{2,k}\ \beta_m + \epsilon_k \quad \text{for } k = 1, \cdots, N_d. \qquad (14)
$$

With BOLS, each new observation allows to update the parameters $\beta_{y,k}$ and $\beta_{m,k}$.[18] In our approach, we re-initialize the estimation procedure for each new day. For this reason, we now distinguish the estimation performed for the first day from the subsequent ones.

For the first day, we initialize hyper-parameters as $b_{y_0} = b_{m_0} = 0$ and we set $B_{y_0} = I_{n_1}$ and $B_{m_0} = I_{n_2}$, where n_1 and n_2 represent the number of variables in Z_1 and Z_2, respectively. We also define

$$b_0 = \begin{bmatrix} b_{y_0} \\ b_{m_0} \end{bmatrix}, \quad B = \begin{bmatrix} B_{y_0} & 0 \\ 0 & B_{m_0} \end{bmatrix}, \tag{15}$$

and, for trade $k = 2, \cdots, N_d$,

$$\mathcal{Z}_{2:k} = \begin{bmatrix} Z_{1,2} - Z_{1,1} & Z_{2,2} \\ Z_{1,3} - Z_{1,2} & Z_{2,1} \\ \vdots & \vdots \\ Z_{1,k} - Z_{1,k-1} & Z_{2,k} \end{bmatrix} \quad \text{and} \quad \Delta\mathcal{Y}_{2:k} = \begin{bmatrix} y_2 - y_1 \\ y_3 - y_2 \\ \vdots \\ y_k - y_{k-1} \end{bmatrix}.$$

As the price y_k is made available, the parameters are updated as:

$$\begin{aligned} \hat{\beta}_k = \left[\hat{\beta}'_{y,k}, \hat{\beta}'_{m,k}\right]' &= \left[B^{-1} + \mathcal{Z}'_{2:k}\mathcal{Z}_{2:k}\right]^{-1}\left[B^{-1}b_0 + \mathcal{Z}'_{2:k}\Delta\mathcal{Y}_{2:k}\right] \\ &= \left[S_{Z'Z;k-1} + \mathcal{Z}'_{k:k}\mathcal{Z}_{k:k}\right]^{-1}\left[S_{Z'Y;k-1} + \mathcal{Z}'_{k:k}\Delta\mathcal{Y}_{k:k}\right], \end{aligned} \tag{16}$$

where $S_{Z'Z;k-1} \equiv B^{-1} + \mathcal{Z}'_{2:k-1}\mathcal{Z}_{2:k-1}$ and $S_{Z'\Delta Y;k-1} \equiv B^{-1}b_0 + \mathcal{Z}'_{2:k-1}\Delta\mathcal{Y}_{2:k-1}$ are sufficient statistics for the parameter estimates β_k, which can be updated with each new price observation y_k. At the end of each day, we obtain $\hat{\beta}_{N_d}$.

For subsequent days, we initialize the hyper-parameters with $b_0 = \hat{\beta}_{N_d}$, meaning that we start the day by defining as hyper-parameters the parameters we obtained at the close of the previous day. Furthermore, we set $B = \left(\frac{n_0}{N_d} S_{Z'Z;t_{N_d}}\right)^{-1}$. Then, we proceed updating the parameters as in Equation (16).

[18]It is easy to verify that the error term ϵ_k is defined as $\epsilon_k = \sigma_{y,k}\ z_{y,k} - \sigma_{y,k-1}\ z_{y,k-1} + \sigma_{m,k}\ z_{m,k}$. The OLS estimator is still consistent in this context.

6.1.2 Estimating State and Variances via Particle Filter

At this stage, we have described how to obtain estimates of model's parameters. Now we describe how the state (m_k) and the variances of the continuous shocks ($\sigma^2_{y,k}$ and $\sigma^2_{m,k}$) are estimated via particle learning, as described by Carvalho et al. (2010). As we can estimate the parameters β_y and β_m via BOLS, this approach appears to be the most efficient way according to the simulation experiment reported by Lopes and Tsay (2011). For the time being, we assume that no jump occurred.

Estimation of the variances.

We distinguish again the first day from the subsequent ones. For the first day, we assume Bayesian priors, as in Lopes and Tsay (2011):

$$m_0 \sim N(\mu_0, c_0), \tag{17}$$

$$\beta_y \sim N(b_{y0}, \sigma^2_y B_{y0}), \qquad \beta_m \sim N(b_{m_0}, \sigma^2_m B_{m_0}), \tag{18}$$

$$\sigma^2_y \sim IG\left(\frac{n_0}{2}, \frac{n_0}{2}\sigma^2_{y0}\right), \qquad \sigma^2_m \sim IG\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma^2_{m_0}\right). \tag{19}$$

In Bayesian literature, it is common to assume that variances follow an inverse-gamma distribution, IG, as it is a natural conjugate prior for the normal distribution.[19]

With each new observation, after estimation of the state m_k, denoted by $\hat{m}_k$, we update the following sufficient statistics of sum of squared residuals:

$$SSR_{y,0} = n_0 \sigma^2_{y0}, \tag{20}$$

$$SSR_{y,k} = SSR_{y,k-1} + (y_k - \hat{x}_k - Z_{1,k}\ \hat{\beta}_{y,k})^2, \tag{21}$$

$$SSR_{m,0} = \nu_0 \sigma^2_{m_0}, \tag{22}$$

$$SSR_{m,k} = SSR_{m,k-1} + (\hat{m}_k - \hat{m}_{k-1} - Z_{2,k}\ \hat{\beta}_{m,k})^2. \tag{23}$$

[19] In practice, to initialize the algorithm, we set $\mu_0 = y_1$ the first log-price in the sample and we set $c_0 = 2 \times \widehat{V}[y_{1:100}]$, where $\widehat{V}[y_{1:100}]$ denotes the estimate of the variance based on the first 100 observations. Following Lopes and Tsay (2011), we also set $n_0 = \nu_0 = 10$. In addition, we set $\sigma^2_{y_0} = 5\ \sigma^2_{y,KF}$ and $\sigma^2_{m_0} = 5\ \sigma^2_{m,KF}$, where $\sigma^2_{y,KF}$ and $\sigma^2_{m,KF}$ are the estimates of the innovation variances obtained from the Kalman filter. We also used other scalings for the variances but our eventual estimates were not changed.

By defining $n_k = n_{k-1} + 1$ and $\nu_k = \nu_{k-1} + 1$, we note that resampling standard deviations can be done by drawing from an inverse-gamma distribution:

$$\sigma^2_{y,k} \sim IG\left(\frac{n_k}{2}, \frac{1}{2}SSR_{y,k}\right) \quad \text{and} \quad \sigma^2_{m,k} \sim IG\left(\frac{\nu_k}{2}, \frac{1}{2}SSR_{m,k}\right).$$

For each new day, we re-initialize the SSR with:

$$SSR_{y,0} = \frac{n_0}{N_d}SSR_{y,N_d} \quad \text{and} \quad SSR_{m,0} = \frac{\nu_0}{N_d}SSR_{m,N_d}.$$

The idea of doing so is that the best parameter estimate as the market opens is yesterday's close, although the uncertainty surrounding this observation might be large. As the new day evolves, parameter estimates will adjust to new values and the variances (filtered for intradaily seasonality) will in principle decrease.

Estimation of the fundamental value with known parameters.

To cast our model within the particle filter literature, we notice that Equations (6) and (7) can be rewritten as, for $k = 1, 2, \cdots, N_d$:

$$y_k | m_k, Z_k \sim p(y_k | m_k, Z_k), \tag{24}$$

$$m_k | m_{k-1}, Z_k \sim p(m_k | m_{k-1}, Z_k). \tag{25}$$

We have regrouped all predetermined variables in a single vector Z_k. We denote by p a generic probabilistic model that needs to be specified depending on the particular problem.[20] If parameters were known, two fundamental approaches could be used to estimate the latent state m_k.

The seminal approach, due to Gordon, Salmond, and Smith (1993), called **Bootstrap Filter**, proceeds as follows:

1. At the initial step 0, simulate M particles $m_0^{(i)} \sim N(\mu_0, c_0)$, for $i = 1, \cdots, M$.

[20] This general notation allows for a potentially non-linear and non-Gaussian model. Even though our model is linear, we use the particle filter, as transaction-price innovation is not Gaussian. In addition, particle filter is a convenient setting to update parameter estimates with each new observation.

2. At step k, propagate the particle $m_{k-1}^{(i)}$ to $\tilde{m}_k^{(i)}$ using $p(m_k|m_{k-1}, Z_k)$ (Equation (25)).

3. Resample $m_k^{(i)}$ from candidate particles $\tilde{m}_k^{(i)}$ by drawing with resampling, where particle $m_k^{(i)}$ is chosen with a probability proportional to the weight $w_k^{(i)} \propto p(y_k|\tilde{m}_k^{(i)}, Z_k)$.

Having described this algorithm, several remarks are of order. First, in step (2), we propagate $m_{k-1}^{(i)}$ to $\tilde{m}_k^{(i)}$ by using:

$$\tilde{m}_k^{(i)} = m_{k-1}^{(i)} + Z_{2,k}\ \beta_{m,k} + \sigma_{m,k}\ z_{m,k}^{(i)},$$

where $z_{m,k}^{(i)}$ is drawn from a Gaussian $N(0,1)$. In other words, we do not allow for jumps here. The reason for this is that we want to obtain a conservative value of $\tilde{m}_k^{(i)}$, which, when confronted with y_k, will allow us to detect if an abnormal realization of y_k took place. And, indeed, a first way to detect jumps is to consider the likelihood $p(y_k|\tilde{m}_k^{(i)}, Z_k)$ for all the candidate particles. There are cases where, even for a very large amount of particles, M, all the likelihoods are infinitesimally small. Such cases would clearly qualify as jumps given that the observations just do not match the model.

Second, if no jump is detected, meaning that likelihoods $p(y_k|\tilde{m}_k^{(i)}, Z_k)$ are not all infinitesimally small, it is still possible that the realization of y_k be highly unlikely given the current parameter estimates and m_k. To investigate this issue, we construct the predictive distribution $p(y_k|y^{k-1}, Z^k)$, where $y^k = \{y_k, y_{k-1}, \cdots, y_1\}$ and $Z^k = \{Z_k, Z_{k-1}, \cdots, Z_1\}$, and consider if the actual observation y_k can come from this posterior distribution with reasonable probability.[21]

To obtain this predictive distribution, we follow the approach described by Maiz, Miguez, and Djuric (2009). The predictive density is defined as:

$$p(y_k|y^{k-1}, Z^k) = \int p(y_k|m_k, Z^k) p(m_k|y^{k-1}, Z^k) dm_k. \tag{26}$$

[21] We always assume in determining the predictive distribution that the explanatory variables of the model are known. In practice, as the time of the next trade k and the traded price y_k become known, the other right-hand-side variables for our model also become known.

To simulate from this density, it is necessary to sample from $p(m_k|y^{k-1}, Z^k)$, defined as:

$$p(m_k|y^{k-1}, Z^k) = \int p(m_k|m_{k-1}, Z^k)p(m_{k-1}|y^{k-1}, Z^k)dm_{k-1}$$
$$\approx \frac{1}{M}\sum_{i=1}^{M} p(m_k|m_{k-1}^{(i)}, Z^k). \tag{27}$$

The reason for this is that the particles resulting from the boostrap filter provide a sample representation of $p(m_{k-1}|y^{k-1}, Z^k)$, see Gordon, Salmond, and Smith (1993, p. 108). Contemplating Equation (27), we note that the predictive density can be reinterpreted as a mixture of distributions, from which it is easy to sample. The algorithm is now traced. We start with simulating from Equation (27) a sample of $i' = 1, \cdots, M'$ draws. To do so, we uniformly draw from the particles $m_{k-1}^{(i)}$ and for each draw we generate $\tilde{m}_k^{(i')}$ using Equation (25). This yields a sample drawn from $p(m_k|y^{k-1}, Z^k)$.

Then, as a next step, we observe that Equation (26) can be approximated as:

$$p(y_k|y^{k-1}, Z^k) \approx \frac{1}{M'}\sum_{i'=1}^{M'} p(y_k|\tilde{m}_k^{(i')}, Z^k).$$

Thus, the density of y_k is a mixture of distributions from which we can easily sample. We consider M'' draws obtained as:

$$y_k^{(i'')} = m_k^{(i'')} + Z_{1,k}\ \beta_{y,k} + \sigma_{y,k}\ z_{y,k}^{(i'')}, \quad \text{for } i'' = 1, \cdots, M''. \tag{28}$$

where the $m_k^{(i'')}$ are redrawn among the $\tilde{m}_k^{(i')}$ (Step 3). The $y_k^{(i'')}$ constitute a sample drawn from the posterior distribution that will be used for jump detection.

Even though the Bootstrap Filter, as explained above, plays a crucial role in the detection of jumps, it turns out that for the actual parameter estimation the so-called **Auxiliary Particle Filter** (APF) of Pitt and Shephard (1999) plays a particular role. Whereas the Bootstrap Filter starts by propagating and then resampling, the APF is somewhat more efficient, as it avoids us to throw away some of the resampled $\tilde{m}_k^{(i)}$. This algorithm is based on the following steps, where we follow Lopes and Tsay (2011):

1. Resample $\tilde{m}_{k-1}^{(i)}$ from $m_{k-1}^{(i)}$ using as weights $w_k^{(i)} \propto p(y_k|g(m_{k-1}^{(i)}), Z_k)$.

2. Propagate $\tilde{m}_{k-1}^{(i)}$ to $\tilde{m}_k^{(i)}$ using $p(m_k|\tilde{m}_{k-1}, Z_k)$.

3. Resample $m_k^{(i)}$ from $\tilde{m}_k^{(i)}$ with weights $w_k^{(i)} \propto p(y_k|\tilde{m}_k^{(i)}, Z_k)/p(y_k|g(\tilde{m}_{k-1}^{(i)}), Z_k)$.

In the first step of this algorithm, function g denotes for instance the expected value of m_k:

$$g(m_{k-1}^{(i)}) = E_{k-1}\left[m_k\right] = m_{k-1}^{(i)} + Z_{2,k}\ \beta_{m,k}.$$

This implies that in the second step we use particles $\tilde{m}_{k-1}^{(i)}$ that are of relevance for y_k. Because of this, the algorithm is generally more efficient for the estimation of the latent state and the parameters.

Even though the algorithm is more efficient for parameter estimation, it is less adapted in the case where y_k incorporates a jump. Indeed, if a permanent jump took place at trade $k-1$, then m_k will have adjusted. This is not taken into account in the APF approach as only $g(m_{k-1}^{(i)})$ is used. For this reason, we proceed in two steps. First, we use the Bootstrap Filter to detect jumps and then use an algorithm involving APF for parameter estimation.

Estimation of the fundamental value with unknown parameters.

So far, we assumed the parameters to be known. We now consider the situation where the parameters have to be estimated. For this purpose, we rely on the **Particle Learning** (PL) algorithm of Carvalho et al. (2010). This method requires that parameters can be estimated from sufficient statistics. Other algorithms for parameter estimation, such as Storvik (2002), similarly require that parameters can be updated by using sufficient statistics.

We denote by $s_k = \mathcal{S}(s_{k-1}, m_k, y_k, Z_k)$ and by $s_k^m = \mathcal{K}(s_{k-1}^m, \theta, y_k, Z_k)$ the parameter- and state-sufficient statistics. The PL algorithm is given by the following steps, adapted from Lopes and Tsay (2011):

1. Resample $(\tilde{\theta}, \tilde{s}_{k-1}^m, \tilde{s}_{k-1})$ from $(\theta, s_{k-1}^m, s_{k-1})$ with weights $w_k \propto p(y_k|s_{k-1}^m, \theta, Z_k)$.

2. Sample m_k from $p(m_k|\tilde{s}_{k-1}^m, \tilde{\theta}, y^k, Z^k)$.

3. Update parameter-sufficient statistics: $s_k = \mathcal{S}(\tilde{s}_{k-1}, m_k, y_k, Z_k)$.

4. Sample θ from $p(\theta|s_k)$.

5. Update state-sufficient statistics: $s_k^m = \mathcal{K}(\tilde{s}_{k-1}^m, \theta, y_k, Z_k)$.

6.1.3 Complete Algorithm

For our model, we have already seen how sufficient statistics can be obtained for the estimation of $\beta_{y,k}$ and $\beta_{m,k}$ during the day as a new y_k becomes available. For the problem at hand, we adapted the PL algorithm as follows:

Step 1: Initialization of the algorithm.

a) Simulate $i = 1, \cdots, M$ particles for the initial state $m_0^{(i)} \sim N(\mu_0, c_0)$.

b) Simulate $i = 1, \cdots, M$ particles for the variances $\sigma_{y,1}^{2(i)} \sim IG(\frac{n_0}{2}, \frac{n_0}{2}\sigma_{y_0}^2)$ and $\sigma_{m,1}^{2(i)} \sim IG(\frac{\nu_0}{2}, \frac{\nu_0}{2}\sigma_{m_0}^2)$.

Step 2: Jump detection (Bootstrap filter). In presence of a jump at trade k, all the statistics would be biased if they were computed using price y_k. The objective is therefore to construct the prediction distribution of y_k without using y_k and to test for the occurrence of a jump.

a) Compute $\bar{\sigma}_{y,k}^2$ and $\bar{\sigma}_{m,k}^2$ the variances of the observation and state equations by averaging over the M particles.

b) Propagate $m_{k-1}^{(i)}$ to $\tilde{m}_k^{(i)}$ as:

$$\tilde{m}_k^{(i)} = m_{k-1}^{(i)} + Z_{2,k}\ \beta_{m,k} + \bar{\sigma}_{m,k}\ z_{m,k}^{(i)}$$

with $z_{m,k}^{(i)}$ drawn from the $N(0,1)$ distribution (step (2) of the Bootstrap filter).

c) Resample $m_k^{(i)}$ from particle $\tilde{m}_k^{(i)}$ using weights proportional to $p(y_k|\tilde{m}_k^{(i)}, Z_k)$ (step (3) of the Bootstrap filter).

d) Simulate $y_k^{(i)}$ using:

$$y_k^{(i)} = Z_{1,k}\ \beta_{y,k} + m_k^{(i)} + \bar{\sigma}_{y,k}\ z_{y,k}^{(i)}.$$

Given the non-Gaussianity of the transaction-price innovation (even after removing large jumps), we allow $z_{y,k}^{(i)}$ to be drawn from a Student t distribution with a degree of freedom equal to the inverse of the tail index estimated over the data.[22] We construct the predictive distribution for y_k and compute the quantiles at 0.1% and 99.9%, denoted by $q_y^{0.1\%}$ and $q_y^{99.9\%}$, respectively. These quantiles are used to detect whether a jump occurred at trade k and whether it is a negative or positive jump.

e) To differentiate between permanent and transitory jumps, we make the following assumption: a transitory jump lasts only one trade, whereas a permanent jump lasts at least two trades. Therefore, after trade k, we can test whether trade $k-1$ was a transitory or permanent jump:

- If $q_y^{0.1\%} < y_{k-1} < q_y^{99.9\%}$, we conclude that y_{k-1} was not a jump and go to Step 3.
- If ($y_{k-1} < q_y^{0.1\%} < y_k$) or if ($y_k < q_y^{99.9\%} < y_{k-1}$), we conclude that y_{k-1} was a transitory jump. We do not change the fundamental value m_{k-1} and keep the same predictive distribution. We then go to next trade $k+1$.
- If ($y_{k-1} < q_y^{0.1\%}$ and $y_k < q_y^{0.1\%}$) or if ($y_{k-1} > q_y^{99.9\%}$ and $y_k > q_y^{99.9\%}$), we conclude that y_{k-1} was a permanent jump. We change the fundamental value $m_k = y_{k-1}$ and go back to Step 2 with trade k to update the predictive distribution. We then go to the next trade $k+1$.

Step 3: Parameter estimation (Auxiliary particle filter with Particle learning). If trade k does not correspond to a jump, the statistics can be estimated using price y_k as we know that this is not a jump. If trade k is a transitory jump, we just ignore price y_k.

[22] More precisely, we estimate the left and right tail index (using Hill's estimator with a threshold of 5% of observations). As the skewness of most of the security returns is close to 0, we take the average of the left and right tail indices, which gives us an estimate of the degree of freedom of the t distribution.

If trade k is a permanent jump, we have updated the fundamental value to the previous price (Step 2). We now update the parameters, the variances, and the fundamental value for trade k.

a) Update parameters $\beta_{y,k}$ and $\beta_{m,k}$, as described in Section 6.1.1.

b) Simulate the price process, conditional on parameters at trade k and on the fundamental value at trade $k-1$:

$$y_k^{(i)} = Z_{1,k}\ \beta_{y,k} + m_{k-1}^{(i)} + Z_{2,k}\ \beta_{m,k} + \sigma_{y,k}^{(i)}\ z_{y,k} + \sigma_{m,k}^{(i)}\ z_{m,k}.$$

We denote by $l^{(i)}$ the likelihood of particle $y_k^{(i)}$ conditional on its mean $Z_{1,k}\ \beta_{y,k} + m_{k-1}^{(i)} + Z_{2,k}\ \beta_{m,k}$ and its variance $\sigma_{y,k}^{(i)2} + \sigma_{m,k}^{(i)2}$.

c) Resample from the sufficient statistics ($\sigma_{y,k}^{(i)}$, $\sigma_{m,k}^{(i)}$, and $m_{k-1}^{(i)}$) using as weights: $w^{(i)} = l^{(i)} / \sum_{i=1}^{M} l^{(i)}$. This gives us $\tilde{\sigma}_{y,k}^{(i)}$, $\tilde{\sigma}_{m,k}^{(i)}$, and $\tilde{m}_{k-1}^{(i)}$.

d) Propagate the fundamental value $\tilde{m}_{k-1}^{(i)}$ to $m_k^{(i)}$, using:

$$m_k^{(i)} = \mu^{(i)} + \sqrt{V^{(i)}}\ z_{m,k}^{(i)},$$

where the best predictor of the mean is:

$$\mu^{(i)} = V^{(i)} \times \left(\frac{y_k - Z_{1,k}\beta_{y,k}}{\tilde{\sigma}_{y,k}^{2(i)}} + \frac{\tilde{m}_{k-1}^{(i)}}{\tilde{\sigma}_{m,k}^{2(i)}} \right),$$

and the precision for each particle is:

$$1/V^{(i)} = 1/\tilde{\sigma}_{y,k}^{2(i)} + 1/\tilde{\sigma}_{m,k}^{2(i)},$$

and with $z_{m,k}^{(i)}$ drawn from the $N(0,1)$ distribution.

e) Update sufficient statistics as in Equations (21) and (23) for all particles:

$$\begin{aligned} n_k &= n_{k-1} + 1, \\ SSR_{y,k}^{(i)} &= SSR_{y,k-1}^{(i)} + (y_k - m_k^{(i)} - Z_{1,k}\ \beta_{y,k})^2, \\ \nu_k &= \nu_{k-1} + 1, \\ SSR_{m,k}^{(i)} &= SSR_{m,k-1}^{(i)} + (m_k^{(i)} - m_{k-1}^{(i)} - Z_{2,k}\ \beta_{m,k})^2. \end{aligned}$$

and generate new particles for the innovation and observation error variances of the next trade:

$$\begin{aligned} \sigma_{y,k+1}^{2(i)} &\sim IG\left(\frac{n_k}{2}, \frac{1}{2} SSR_{y,k}^{(i)}\right), \\ \sigma_{m,k+1}^{2(i)} &\sim IG\left(\frac{\nu_k}{2}, \frac{1}{2} SSR_{m,k}^{(i)}\right). \end{aligned}$$

With this last step, we go back to Step 2 with the next observation y_{k+1}.

6.2 Appendix 2: Intradaily Periodic Volatility

Different approaches have been used in the literature to deal with intradaily periodic volatility patterns. Some authors have ignored this issue (Duan and Fulop, 2007), others have estimated their model over arbitrary 30 minute time intervals (MRR, 1997). Still others include the estimation of this component within the general setting of their model (Engle and Sun, 2007).

In this section, we build on Boudt, Croux, and Laurent (2011). Their approach recognize first the possibility that volatility can change from day to day, this is the daily volatility component. They remove this component in a preliminary step. Since intraday returns could contain jumps, this daily volatility should be estimated in a manner which is robust to jumps, which can be achieved by using, for instance, bipower variation.

Specifically, denote by m the intraday sampling frequency, here chosen to be 10 minutes. Denote by $p_{d,im}$ the price that is closest to the $i \cdot m$ th minute on day d. We have $i = 1, \cdots, M$. Let the m-minute log-returns be $r_{d,im} \equiv p_{d,im} - p_{d,(i-1)m}$. The realized bipower variation for day d is:

$$RBV_d \equiv \mu_1^{-2} \sum_{i=2}^{M} |r_{d,im}||r_{d,(i-1)m}|,$$

where $\mu_1 \equiv \sqrt{2/\pi} \simeq 0.798$ under normality, and intradaily standardized returns are then defined as:

$$\bar{r}_{d,im} = \frac{r_{d,im}}{\sqrt{RBV_d}}.$$

At this stage, we could proceed computing a standard deviation using the m-minute returns over several days, as in Taylor and Xu (1997). However, as shown in Boudt, Croux, and Laurent (2008), such a procedure is not appropriate in the presence of jumps in the data generating process, as standard deviation is not a robust estimator. In this context, a more appropriate approach consists in using a scale measure from the robust statistics literature, as in Boudt, Croux, and Laurent (2011).

This latter approach involves the Shortest-Half-Scale (SHS) estimator of Rousseeuw and Leroy (1988). The SHS is an equivalent measure to standard deviation, however, it is outlier-robust. To compute the SHS estimator, we first need to rank returns by size. In the following, n_i denotes the number of sample observations for intraday period i and $\{\bar{r}_{l;i}\}_{l=1,\cdots,n_i}$ is the sample of observations for this intraday period i. We obtain the order statistics $\bar{r}_{(1);i} \leq \bar{r}_{(2);i} \leq \ldots \leq \bar{r}_{(n_i);i}$. Halves length of $h_i = \lfloor n_i/2 \rfloor + 1$ contiguous order observations are defined as $\bar{r}_{(h_i);i} - \bar{r}_{(1);i}$, …, $\bar{r}_{(n_i);i} - \bar{r}_{(h_i-1);i}$, respectively. The shortest half scale is the smallest length of all "halves length" corrected for consistency under normality:

$$\text{ShortH}_i = 0.741 \ \min\{\bar{r}_{(h_i);i} - \bar{r}_{(1);i}, \ldots, \bar{r}_{(n_i);i} - \bar{r}_{(h_i-1);i}\}.$$

Finally, we consider:

$$\hat{\delta}_i^{\text{ShortH}} = \frac{\text{ShortH}_i}{\sqrt{\frac{1}{M}\sum_{j=1}^{M} \text{ShortH}_j^2}}, \tag{29}$$

whose average of squares is equal to one. Eventually, Boudt, Croux, and Laurent (2011) propose the use of the so-called Weighted Standard Deviation (WSD) as the intradaily volatility estimator. The WSD can now be computed for each intraday period across sample days. This estimator is a robust scale estimator that we use as a proxy for intradaily volatility. It is defined as:

$$\hat{\delta}_i^{\text{WSD}} = \frac{\text{WSD}_i}{\sqrt{\frac{1}{M}\sum_{j=1}^{M}\text{WSD}_j^2}},$$

where

$$\text{WSD}_j = \sqrt{1.081\,\frac{\sum_{l=1}^{n_j} w[(\overline{r}_{l;j}/\hat{\delta}_j^{\text{ShortH}})^2]\overline{r}_{l;j}^2}{\sum_{l=1}^{n_j} w[(\overline{r}_{l;j}/\hat{\delta}_j^{\text{ShortH}})^2]}}. \tag{30}$$

The function $w(\cdot)$ in Equation (30) robustifies the standard deviation. It is an indicator equal to one when its argument can not be rejected to be a realization from a $\chi^2(1)$ distribution for a given level of probability, and zero otherwise. In our numerical implementation, $w(z)$ is equal to one when $z \leq 6.635$, which is the 99th percentile of the $\chi^2(1)$. As noted in Boudt, Croux, and Laurent (2011), the SHS estimator is highly robust to jumps, but it has only 37% efficiency with normally distributed $\overline{r}_{d,i}$, against 69% for the WSD. This justifies why the latter is preferred over the former in the paper.

References

[1] Aït-Sahalia, Y., and J. Jacod (2009): Testing for Jumps in a Discretely Observed Process, Annals of Statistics, 37, 422–457.

[2] Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005): How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, Review of Financial Studies, 18, 351–416.

[3] Amihud, Y., and H. Mendelson (1986): Asset Pricing and the Bid-Ask Spread, Journal of Financial Economics, 17, 223–249.

[4] Andersen, T. G., and T. Bollerslev (1997): Intraday Periodicity and Volatility Persistence in Financial Markets, Journal of Empirical Finance, 4, 115–158.

[5] Andersen, T. G., and T. Bollerslev (1998): DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements and Longer Run Dependencies, Journal of Finance, 53, 219–265.

[6] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): The Distribution of Realized Exchange Rate Volatility, Journal of the American Statistical Association, 96, 42–55.

[7] Ball, C. A., and W. N. Torous (1985): On Jumps in Common Stock Prices and their Impact on Call Option Pricing, Journal of Finance, 40, 155–173.

[8] Bandi, F. M., and J. R. Russel (2008): Microstructure Noise, Realized Variance, and Optimal Sampling, Review of Economic Studies, 75, 339–369.

[9] Barndorff-Nielsen, O., and N. Shephard (2002): Estimating Quadratic Variation Using Realized Variance, Journal of Applied Econometrics, 17, 457–478.

[10] Barndorff-Nielsen, O., and N. Shephard (2004): Power and Bipower Variation with Stochastic Volatility and Jumps, Journal of Financial Econometrics, 2, 1–37.

[11] Barndorff-Nielsen, O., and N. Shephard (2006): Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation, Journal of Financial Econometrics, 4, 1–30.

[12] Biais, B., P. Hillion, and C. Spatt (1995): An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse, Journal of Finance, 50, 1655–1689.

[13] Bollerslev, T., T. H. Law, and G. Tauchen (2008): Risk, Jumps, and Diversification, Journal of Econometrics, 144, 234–256.

[14] Bos, C. (2008): Model-based Estimation of High-frequency Jump Diffusions with Microstructure Noise and Stochastic Volatility, Tinbergen Institute Discussion Paper No. TI 2008-011/4.

[15] Boudt, K., C. Croux, and S. Laurent (2011): Robust Estimation of Intraweek Periodicity in Volatility and Jump Detection, Journal of Empirical Finance, 18, 353-367.

[16] Brennan, M. J., and A. Subrahmanyam (1996): Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns, Journal of Financial Economics, 41, 441–464.

[17] Carvalho, C. M., M. Johannes, H. F. Lopes, and N. Polson (2010): Particle Learning and Smoothing, Statistical Science, 25, 88–106.

[18] Chan, L. K. C., and J. Lakonishok (1993): Institutional Trades and Intra-day Stock Price Behavior, Journal of Financial Economics, 33, 173–200.

[19] Christensen, K., R. C. A. Oomen, and M. Podolskij (2014): Fact or Friction: Jumps at Ultra High Frequency, Journal of Financial Economics, 114, 576–599.

[20] De Jong, F., T. Nijman, and A. Röell (1996): Price Effects of Trading and Components of the Bid-ask Spread on the Paris Bourse, Journal of Empirical Finance, 3, 193–213.

[21] Diamond, D. W., and R. E. Verrecchia (1987): Constraints on Short-Selling and Asset Price Adjustment to Private Information, Journal of Financial Economics, 18, 277–311.

[22] Diebold, F. X., and G. Strasser (2013): On the Correlation Structure of Microstructure Noise: A Financial Economic Approach. Review of Economic Studies, 80, 1304–1337.

[23] Doucet, A., and A. M. Johansen (2009): A Tutorial on Particle Filter and Smoothing: Fifteen Years Later, in *The Oxford Handbook of Nonlinear Filtering,* ed. by D. Crisan, and B. Rozovsky, chap. 24, pp. 656–704. Oxford University Press.

[24] Duan, J. C., and A. Fulop (2007): How Frequently Does the Stock Price Jump? An Analysis of High-Frequency Data with Microstructure Noises, Magyar Nemzeti Bank Working Paper No. 2007/4.

[25] Dufour, A., and R. F. Engle (2000): Time and the Price Impact of a Trade, Journal of Finance, 55, 2467–2498.

[26] Easley, D., and M. O'Hara (1992): Time and the Process of Security Price Adjustment, Journal of Finance, 47, 577–605.

[27] Ellis, K., R. Michaely, and M. O'Hara (2000): The Accuracy of Trade Classification Rules: Evidence from Nasdaq, Journal of Financial and Quantitative Analysis, 35, 529–551.

[28] Engle, R. F., and J. R. Russell (1998): Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, Econometrica, 66, 1127–1162.

[29] Engle, R. F., and Z. Sun (2007): When Is Noise not Noise - A Microstructure Estimate of Realized Volatility, NYU Working Paper No. FIN-07-047.

[30] Foster, F. D., and S. Viswanathan (1993): Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models, Journal of Finance, 48, 187–211.

[31] Foucault, T., S. Moinas, and E. Theissen (2007): Does Anonymity Matter in Electronic Limit Order Markets?, Review of Financial Studies, 5, 1707–1747.

[32] Fox, A. J. (1972): Outliers in Time Series, Journal of the Royal Statistical Society, Series B, 34, 350–363.

[33] Frijns, B., and P. C. Schotman (2004): Price Discovery in Tick Time, CEPR Discussion Paper No. 4456.

[34] Glosten, L. R., and L. Harris (1988): Estimating the Components of the Bid-Ask Spread, Journal of Financial Economics, 21, 123–142.

[35] Glosten, L. R., and P. R. Milgrom (1985): Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, Journal of Financial Economics, 14, 71–100.

[36] Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993): Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation, IEE Proceedings-F, 140, 107–113.

[37] Hameed, A., W. Kang, and S. Viswanathan (2010): Stock Market Declines and Liquidity, Journal of Finance, 65, 257–293.

[38] Hasbrouck, J. (1991): Measuring the Information Content of Stock Trades, Journal of Finance, 46, 179–207.

[39] Holthausen, R. W., R. W. Leftwich, and D. Mayers (1987): The Effect of Large Block Transactions on Security Prices: A Cross-sectional Analysis, Journal of Financial Economics, 19, 237–268.

[40] Jain, P. J., and G. Joh (1988): The Dependence between Hourly Prices and Trading Volume, Journal of Financial and Quantitative Analysis, 23, 269–283.

[41] Johannes, M. S., N. G. Polson, and J. R. Stroud (2009): Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices, Review of Financial Studies, 22, 2259–2299.

[42] Keim, D. B., and A. Madhavan (1995): Anatomy of the Trading Process: Empirical Evidence on the Behavior of Institutional Traders, Journal of Financial Economics, 37, 371–398.

[43] Kitagawa, G. (1998): A Self-Organizing State-Space Model, Journal of the American Statistical Association, 93(443), 1203–1215.

[44] Kyle, A. (1985): Continuous Auctions and Insider Trading, Econometrica, 53, 1315–1335.

[45] Lahaye, J., S. Laurent, and C. J. Neely (2011), Jumps, Cojumps and Macro Announcements, Journal of Applied Econometrics 6, 893–921.

[46] Lee, C., and M. Ready (1991): Inferring Trade Direction from Intraday Data, Journal of Finance, 46, 733–746.

[47] Lopes, H. F., and R. S. Tsay (2011): Particle Filters and Bayesian Inference in Financial Econometrics, Journal of Forecasting, 30, 168–209.

[48] Madhavan, A., M. Richardson, and M. Roomans (1997): Why Do Security Prices Change? A Transaction-level Analysis of NYSE Stocks, Review of Financial Studies, 10, 1035–1064.

[49] Madhavan, A., and S. Smidt (1991): A Bayesian Model of Intraday Specialist Pricing, Journal of Financial Economics, 30, 99–134.

[50] Maheu, J. M., and T. H. McCurdy (2004): News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns, Journal of Finance, 59, 755–793.

[51] Maiz, C. S., J. Miguez, and P. M. Djuric (2009): Particle Filtering in the Presence of Outliers, in 15th Workshop on Statistical Signal Processing, 33–36. IEEE.

[52] Parlour, C. A. (1998): Price Dynamics in Limit Order Markets, Review of Financial Studies, 11, 789–816.

[53] Pitt, M. K., and N. Shephard (1999): Filtering via Simulation: Auxiliary Particle Filters, Journal of the American Statistical Association, 94, 590–599.

[54] Roll, R. (1984): A Simple Implicit Measure of the Effective Bid-ask Spread in an Efficient Market, Journal of Finance, 39, 1127–1139.

[55] Rousseeuw, P., and A. Leroy (1988): A Robust Scale Estimator Based on the Shortest Half, Statistica Neerlandica, 42, 103–116.

[56] Rydberg, T. H., and N. Shephard (2003): Dynamics of Trade-by-Trade Price Movements: Decomposition and Models, Journal of Financial Econometrics, 1, 2–25.

[57] Saar, G. (2001): Price Impact Asymmetry of Block Trades: An Institutional Trading Explanation, Review of Financial Studies, 14, 1153–1181.

[58] Sadka, R. (2006): Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk, Journal of Financial Economics, 80, 309–349.

[59] Storvik, G. (2002): Particle Filters for State-space Models with the Presence of Unknown Static Parameters, IEEE Transactions on Signal Processing, 50, 281–289.

[60] Taylor, S. J., and X. Xu (1997): The Incremental Volatility Information in One Million Foreign Exchange Quotations, Journal of Empirical Finance, 4, 317–340.

[61] Venkataraman, K. (2001): Automated Versus Floor Trading: An Analysis of Execution Costs on the Paris and New York Exchanges, Journal of Finance, 56, 1445–1485.

Table 1: Sample summary statistics

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Market cap. (end of 2002)	5117	1338	22537	19886	5387	19181	31728	3499	16820	17139	93267	16749
Number of trades	187662	47706	140116	169447	35180	77140	66989	28823	115588	114455	144601	119534
Avg trade price (euro)	6.307	4.379	11.912	22.437	36.646	38.208	7.451	23.547	18.040	16.726	125.086	15.630
Avg trade size (000 shares)	4.678	1.753	2.721	1.767	0.533	0.813	3.417	0.756	2.105	1.929	0.884	2.108
Avg trade size (000 euros)	29.539	7.515	32.243	39.521	19.457	31.005	25.528	17.756	37.788	31.910	110.461	32.915
Avg duration (seconds)	6.734	26.465	9.018	7.458	35.803	16.361	18.852	43.507	10.925	11.036	8.740	10.569
Med trade size (000 shares)	1.100	0.608	1.157	0.500	0.200	0.372	0.987	0.300	1.000	0.830	0.432	0.855
Med trade size (000 euros)	7.008	2.727	13.836	10.878	7.565	14.115	7.126	7.148	18.355	13.510	54.160	13.083
Med duration (seconds)	4.000	13.000	4.000	4.000	13.000	6.000	9.000	15.000	4.000	5.000	4.000	5.000
Avg return (bp)	0.016	-0.162	-0.015	0.014	-0.040	-0.006	0.025	-0.009	-0.012	-0.030	-0.006	-0.018
Volatility (bp)	13.135	25.028	10.773	9.218	15.142	9.443	14.575	17.388	7.131	11.921	6.273	11.547
Volatility per time unit (bp)	8.226	10.685	5.812	5.400	5.274	4.208	6.230	5.627	3.788	6.373	3.454	5.746
Skewness	-0.003	0.157	0.058	0.011	0.035	-0.003	0.000	-0.059	0.001	2.451	0.037	-1.405
Kurtosis	5.343	9.715	7.249	8.905	19.383	17.788	7.033	16.644	14.301	135.404	4.476	89.319
Tail index	0.336	0.381	0.326	0.319	0.447	0.422	0.239	0.468	0.311	0.425	0.404	0.360
Zero price changes (%)	0.485	0.504	0.445	0.403	0.397	0.415	0.506	0.407	0.449	0.429	0.527	0.430
Buy-initiated order (%)	0.445	0.573	0.463	0.398	0.440	0.444	0.386	0.492	0.478	0.447	0.451	0.426
Sell-initiated order (%)	0.496	0.409	0.491	0.547	0.532	0.520	0.587	0.484	0.469	0.510	0.507	0.533
Initially unclassified (%)	0.059	0.018	0.046	0.056	0.027	0.036	0.027	0.024	0.053	0.043	0.042	0.041

Note: The table reports summary statistics on the stocks under study. For the stock itself, it reports the market capitalization (in million euros, the number of trades, trade price, trade size (in number of shares and euros), and duration (in seconds). Then for the return process, it reports the average return, volatility, skewness, kurtosis, and Hill's tail index (average of the left and right estimates based on the 5% quantile). It finally reports the proportion of zero price changes and the proportion of buy-initiated and sell-initiated trades.

Table 2: Predictive regression for the signed trading volume

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Constant	0.038	-0.251	-0.009	0.140	-0.013	0.050	0.134	-0.035	-0.037	-0.013	0.039	0.044
	(0.013)	(0.022)	(0.013)	(0.018)	(0.046)	(0.022)	(0.027)	(0.038)	(0.011)	(0.013)	(0.021)	(0.018)
$\sum_{i=k-10}^{k-1} DV_i$	0.268	0.199	0.225	0.214	0.377	0.270	0.301	0.332	0.225	0.246	0.099	0.220
	(0.011)	(0.021)	(0.011)	(0.014)	(0.040)	(0.022)	(0.019)	(0.036)	(0.012)	(0.013)	(0.019)	(0.014)
$\sum_{i=k-10}^{k-1} D_i$	0.164	-0.041	0.102	-0.034	-0.407	-0.384	0.106	-0.398	0.040	-0.095	-0.304	0.012
	(0.017)	(0.027)	(0.015)	(0.024)	(0.085)	(0.037)	(0.031)	(0.067)	(0.015)	(0.018)	(0.032)	(0.021)
$\sum_{i=k-10}^{k-1} r_i$	-0.158	0.581	0.304	1.726	1.138	1.766	0.030	0.987	0.525	0.964	4.608	0.660
	(0.293)	(0.194)	(0.275)	(0.489)	(0.623)	(0.565)	(0.436)	(0.495)	(0.377)	(0.276)	(0.854)	(0.357)
$\sum_{i=k-10}^{k} \tau_i$	0.210	0.237	0.073	0.975	0.122	0.028	0.518	0.062	0.127	0.161	0.194	0.589
	(0.089)	(0.037)	(0.067)	(0.116)	(0.061)	(0.062)	(0.069)	(0.039)	(0.046)	(0.056)	(0.114)	(0.080)
R^2	0.016	0.011	0.011	0.011	0.080	0.053	0.025	0.071	0.009	0.013	0.011	0.010

Note: The table reports the predictive regressions used for the construction of the trade direction and signed trading volume surprises. The signed volume depends on a constant, 10 lags of the signed volume DV_i, 10 lags of the trade direction D_i, 10 lags of the price change r_i, and the current value and 10 lags of the duration since the last trade τ_i. All standard errors, reported in parenthesis, are computed with White's heteroscedasticity correction.

Table 3: Estimation with constant parameters and no jump filtering (Kalman filter)

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Panel A: Model with microstructure variables												
μ	0.132	2.329	0.595	0.816	0.406	0.139	0.614	0.057	0.352	0.288	0.091	1.043
	(0.132)	(0.149)	(0.124)	(0.110)	(0.083)	(0.074)	(0.100)	(0.084)	(0.069)	(0.126)	(0.064)	(0.112)
ϕ^+	1.634	1.990	2.034	2.078	1.642	1.449	2.461	2.137	1.444	2.002	1.597	2.242
	(0.026)	(0.076)	(0.029)	(0.028)	(0.076)	(0.033)	(0.057)	(0.085)	(0.020)	(0.040)	(0.019)	(0.033)
ϕ^-	1.350	6.547	2.110	1.396	1.570	1.321	1.424	2.361	1.688	1.819	1.368	1.879
	(0.021)	(0.161)	(0.026)	(0.019)	(0.052)	(0.027)	(0.036)	(0.093)	(0.022)	(0.029)	(0.015)	(0.025)
λ^+	1.030	2.304	0.921	0.454	0.513	0.310	0.991	0.612	0.395	1.007	0.489	0.731
	(0.019)	(0.099)	(0.025)	(0.016)	(0.054)	(0.027)	(0.039)	(0.075)	(0.018)	(0.046)	(0.011)	(0.025)
λ^-	0.989	2.115	0.932	0.637	0.360	0.395	1.031	0.719	0.495	0.830	0.507	0.827
	(0.017)	(0.107)	(0.023)	(0.013)	(0.047)	(0.024)	(0.028)	(0.073)	(0.020)	(0.030)	(0.011)	(0.021)
$\bar{\phi}^+$	6.440	11.223	3.778	3.076	3.952	3.121	5.505	5.133	1.562	3.872	2.214	3.751
	(0.087)	(0.771)	(0.111)	(0.081)	(0.454)	(0.162)	(0.267)	(0.697)	(0.094)	(0.171)	(0.064)	(0.144)
$\bar{\phi}^-$	7.011	10.409	3.584	3.180	4.649	2.826	7.007	5.177	1.804	3.824	2.255	3.882
	(0.086)	(0.770)	(0.110)	(0.080)	(0.443)	(0.154)	(0.265)	(0.698)	(0.094)	(0.164)	(0.064)	(0.151)
$\bar{\lambda}^+$	-0.995	-2.047	-0.821	-0.530	-0.751	-0.433	-1.079	-0.506	-0.377	-0.908	-0.458	-0.783
	(0.019)	(0.099)	(0.024)	(0.018)	(0.076)	(0.032)	(0.043)	(0.092)	(0.019)	(0.044)	(0.011)	(0.026)
$\bar{\lambda}^-$	-0.990	-1.252	-0.777	-0.737	-0.653	-0.418	-1.397	-0.601	-0.356	-0.568	-0.529	-0.914
	(0.017)	(0.104)	(0.024)	(0.014)	(0.063)	(0.029)	(0.031)	(0.097)	(0.021)	(0.032)	(0.011)	(0.021)
σ_m	5.299	12.750	5.182	4.413	8.447	4.751	6.815	9.312	3.596	6.552	2.874	5.632
	(0.031)	(0.187)	(0.041)	(0.031)	(0.193)	(0.088)	(0.067)	(0.192)	(0.032)	(0.331)	(0.014)	(0.189)
σ_y	20.119	21.968	19.418	17.641	13.799	13.317	13.378	17.397	14.772	22.101	9.435	16.873
	(0.125)	(0.369)	(0.125)	(0.122)	(0.347)	(0.213)	(0.135)	(0.387)	(0.134)	(0.474)	(0.050)	(0.194)
LR_1	114.8	1577.4	10.2	1237.4	9.4	49.6	379.2	3.8	129.0	95.0	221.0	122.4
Panel B: Model without microstructure variables												
σ_m	8.703	17.493	6.637	5.802	9.654	5.588	9.558	10.549	4.235	7.756	4.006	7.112
	(0.028)	(0.165)	(0.040)	(0.031)	(0.190)	(0.083)	(0.066)	(0.198)	(0.033)	(0.281)	(0.016)	(0.161)
σ_y	21.820	24.660	22.146	19.960	15.035	15.013	14.812	19.195	16.221	24.596	11.237	19.172
	(0.124)	(0.366)	(0.130)	(0.127)	(0.347)	(0.197)	(0.146)	(0.390)	(0.136)	(0.399)	(0.056)	(0.185)
LR_2	131416.6	23670.2	55510.4	72707.2	8234.0	22333.4	34272.0	6587.4	30265.2	33235.4	78097.0	45866.0

Note: The table reports the parameter estimates of Equations (6) and (7) defined in the text. Panel A reports the parameter estimates of the complete model, whereas Panel B reports the estimates of the model with no microstructure explanatory variables. All parameters and standard errors have been multiplied by 100. LR_1 is the likelihood ratio statistic for the test of symmetry, i.e., buy and sell orders have the same price impact. It is distributed as a $\chi^2(4)$. The 95% (99%) critical values are 9.49 (13.28). LR_2 is the test of the null that all the parameters but the constant are zero. It is distributed as a $\chi^2(8)$. The 99% critical value is 20.09.

Table 4: Statistics on detected jumps

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Nb of trades	187663	47707	140117	169448	35181	77141	66990	28824	115589	114456	144602	119535
Panel A: Total number of jumps												
J_y^+	265	21	195	185	19	30	138	10	63	65	164	131
J_y^-	254	22	157	143	14	22	94	8	54	53	144	97
J_m^+	184	21	162	144	16	43	72	10	89	65	104	115
J_m^-	210	14	165	154	12	35	101	10	104	69	140	119
Panel B: Average number of jumps per day												
J_y^+	6.310	0.500	4.643	4.405	0.452	0.714	3.286	0.238	1.500	1.548	3.905	3.119
J_y^-	6.048	0.524	3.738	3.405	0.333	0.524	2.238	0.190	1.286	1.262	3.429	2.310
J_m^+	4.381	0.500	3.857	3.429	0.381	1.024	1.714	0.238	2.119	1.548	2.476	2.738
J_m^-	5.000	0.333	3.929	3.667	0.286	0.833	2.405	0.238	2.476	1.643	3.333	2.833
Panel C: Proportion of jumps per day (in % of total number of trades)												
J_y^+	0.141	0.044	0.139	0.109	0.054	0.039	0.206	0.035	0.055	0.057	0.113	0.110
J_y^-	0.135	0.046	0.112	0.084	0.040	0.029	0.140	0.028	0.047	0.046	0.100	0.081
J_m^+	0.098	0.044	0.116	0.085	0.045	0.056	0.107	0.035	0.077	0.057	0.072	0.096
J_m^-	0.112	0.029	0.118	0.091	0.034	0.045	0.151	0.035	0.090	0.060	0.097	0.100
Panel D: Size of the jumps (in % of previous price)												
J_y^+	0.431	1.059	0.430	0.381	0.821	0.596	0.545	1.208	0.330	0.536	0.217	0.484
J_y^-	-0.430	-1.273	-0.426	-0.379	-1.018	-0.552	-0.520	-1.120	-0.314	-0.527	-0.207	-0.472
J_m^+	0.270	0.653	0.306	0.229	0.703	0.461	0.343	1.308	0.304	0.428	0.135	0.357
J_m^-	-0.248	-0.653	-0.296	-0.220	-0.898	-0.435	-0.355	-0.861	-0.292	-0.392	-0.107	-0.314

Note: The table reports for the various stocks information about the jumps of the various types that have been detected. J_y^+ and J_y^- represent positive and negative (transitory) jumps in the observation equations. J_m^+ and J_m^- are positive and negative (permanent) jumps in the state equation. Panel A corresponds to the total number of jumps over 42 days. Panel B displays the average number of jumps per day. Panel C represents the probability that any given observation represents a jump. Panel D represents the average size of the jumps of a given type.

Table 5: Hourly breakdown of the number and frequency of jumps

Hour	J_y^+	J_y^-	J_y	J_m^+	J_m^-	J_m	Total
Panel A: Total number of jumps							
09:00 - 09:59	356	286	642	318	321	639	1281
10:00 - 10:59	167	146	313	109	133	242	555
11:00 - 11:59	113	103	216	64	97	161	377
12:00 - 12:59	74	61	135	44	45	89	224
13:00 - 13:59	61	45	106	43	34	77	183
14:00 - 14:59	90	81	171	67	63	130	301
15:00 - 15:59	128	117	245	74	87	161	406
16:00 - 16:59	192	155	347	137	151	288	635
17:00 - 17:30	105	68	173	55	61	116	289
Panel B: Relative frequency per category							
09:00 - 09:59	27.7	26.9	27.3	34.9	32.4	33.6	30.1
10:00 - 10:59	13.0	13.7	13.3	12.0	13.4	12.7	13.1
11:00 - 11:59	8.8	9.7	9.2	7.0	9.8	8.5	8.9
12:00 - 12:59	5.8	5.7	5.7	4.8	4.5	4.7	5.3
13:00 - 13:59	4.7	4.2	4.5	4.7	3.4	4.0	4.3
14:00 - 14:59	7.0	7.6	7.3	7.4	6.4	6.8	7.1
15:00 - 15:59	10.0	11.0	10.4	8.1	8.8	8.5	9.6
16:00 - 16:59	14.9	14.6	14.8	15.0	15.2	15.1	14.9
17:00 - 17:30	8.2	6.4	7.4	6.0	6.1	6.1	6.8
Panel C: Relative frequency per hour							
09:00 - 09:59	27.8	22.3	50.1	24.8	25.1	49.9	100.0
10:00 - 10:59	30.1	26.3	56.4	19.6	24.0	43.6	100.0
11:00 - 11:59	30.0	27.3	57.3	17.0	25.7	42.7	100.0
12:00 - 12:59	33.0	27.2	60.3	19.6	20.1	39.7	100.0
13:00 - 13:59	33.3	24.6	57.9	23.5	18.6	42.1	100.0
14:00 - 14:59	29.9	26.9	56.8	22.3	20.9	43.2	100.0
15:00 - 15:59	31.5	28.8	60.3	18.2	21.4	39.7	100.0
16:00 - 16:59	30.2	24.4	54.6	21.6	23.8	45.4	100.0
17:00 - 17:30	36.3	23.5	59.9	19.0	21.1	40.1	100.0

Note: The table reports in Panel A the total number of jumps, for the various jump types, for each stock, depending on the time of the day. Panel B reports the relative frequency of jumps for each stock. Formally if N_{idh} denotes the number of jumps found for company i, on day d, and hour h, and if $T_i = \sum_d \sum_h N_{idh}$, then the table presents the relative jump frequency for each hour h defined as the statistics $100 \times \frac{1}{12} \sum_{i=1}^{12} \sum_d \frac{N_{idh}}{T_i}$. The jumps are obtained by using a model with microstructure variables.

Table 6: Estimation with time-varying parameters and jump filtering (particle filter)

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Panel A: Model with microstructure variables												
μ	0.333	0.473	0.164	0.326	0.101	0.140	-0.004	0.103	0.268	0.072	-0.001	0.397
	(0.718)	(1.012)	(0.697)	(0.904)	(0.322)	(0.437)	(0.776)	(0.244)	(0.336)	(0.742)	(0.300)	(0.586)
ϕ^+	2.091	3.317	2.666	2.625	3.307	2.159	3.701	3.316	2.065	2.966	2.051	3.083
	(0.626)	(0.890)	(0.396)	(0.462)	(1.081)	(0.555)	(0.803)	(1.003)	(0.345)	(0.811)	(0.366)	(0.633)
ϕ^-	1.867	6.951	2.652	1.800	2.433	1.968	1.884	3.629	2.326	2.671	1.770	2.448
	(0.540)	(2.383)	(0.501)	(0.390)	(0.765)	(0.496)	(0.637)	(1.145)	(0.396)	(0.915)	(0.302)	(0.742)
λ^+	0.962	1.708	0.889	0.354	0.477	0.311	0.604	0.765	0.423	0.896	0.530	0.628
	(0.275)	(0.809)	(0.283)	(0.195)	(0.419)	(0.356)	(0.261)	(0.469)	(0.208)	(0.569)	(0.138)	(0.283)
λ^-	1.001	1.730	0.854	0.687	0.348	0.484	0.993	0.883	0.507	0.833	0.558	0.928
	(0.306)	(0.774)	(0.245)	(0.184)	(0.360)	(0.298)	(0.243)	(0.505)	(0.248)	(0.347)	(0.127)	(0.395)
$\bar{\phi}^+$	6.716	10.237	3.831	3.180	4.430	3.408	5.534	4.325	1.796	3.743	2.168	3.746
	(1.459)	(5.175)	(1.104)	(0.662)	(2.019)	(1.647)	(1.760)	(3.261)	(0.556)	(1.514)	(0.490)	(1.412)
$\bar{\phi}^-$	7.131	10.984	3.741	3.427	4.361	2.752	6.926	6.163	1.848	3.897	2.169	4.117
	(0.984)	(5.099)	(0.896)	(0.840)	(2.349)	(1.517)	(1.899)	(2.412)	(0.567)	(1.393)	(0.401)	(1.443)
$\bar{\lambda}^+$	-0.901	-1.557	-0.791	-0.468	-0.732	-0.489	-0.820	-0.610	-0.377	-0.795	-0.512	-0.772
	(0.208)	(0.593)	(0.306)	(0.218)	(0.410)	(0.345)	(0.320)	(0.430)	(0.212)	(0.363)	(0.114)	(0.274)
$\bar{\lambda}^-$	-0.923	-1.087	-0.755	-0.737	-0.590	-0.480	-1.341	-0.764	-0.324	-0.653	-0.551	-0.947
	(0.236)	(0.612)	(0.240)	(0.183)	(0.332)	(0.319)	(0.387)	(0.563)	(0.256)	(0.353)	(0.113)	(0.316)
σ_m	5.641	14.464	5.187	4.909	8.223	4.880	6.650	10.019	3.595	13.317	2.904	5.665
	(1.116)	(6.922)	(0.737)	(1.297)	(2.564)	(0.787)	(1.240)	(2.721)	(0.502)	(33.097)	(0.264)	(2.075)
σ_y	15.673	18.441	18.867	13.955	14.430	14.179	14.039	17.287	15.120	16.215	8.831	15.800
	(3.873)	(6.750)	(3.632)	(2.895)	(3.394)	(2.779)	(2.196)	(3.918)	(3.224)	(4.150)	(1.420)	(6.657)
Panel B: Model without microstructure variables												
σ_m	8.047	15.657	6.311	5.721	9.145	6.466	9.517	10.524	4.024	9.338	4.154	7.452
	(2.539)	(5.031)	(2.000)	(1.806)	(2.882)	(2.090)	(3.064)	(3.491)	(1.270)	(3.255)	(1.325)	(2.459)
σ_y	19.920	22.394	21.279	18.919	13.064	14.002	14.704	17.451	13.767	23.653	10.525	24.585
	(6.295)	(7.172)	(6.755)	(5.983)	(4.185)	(4.540)	(4.650)	(5.508)	(4.406)	(7.636)	(3.394)	(9.028)

Note: The table reports average of the parameter estimates and standard deviations of the parameters of the same microstructure model as of Table 3 but using the particle-learning algorithm.

Table 7: Percentage contributions to total return variation

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Contribution of the microstructure model												
Microstructure model	61.689	59.000	47.877	50.257	38.924	40.982	54.488	36.922	36.321	41.963	58.431	45.210
Contribution of the continuous innovations												
η_y	13.554	22.579	13.738	19.984	21.388	14.972	13.411	19.408	11.867	31.443	13.959	16.000
η_m	4.452	4.803	12.996	7.589	17.590	18.103	8.971	18.585	20.346	8.371	8.427	11.558
η_y, η_m	4.969	4.895	8.853	7.489	7.076	8.932	6.914	7.068	8.422	6.192	7.574	8.423
Intraday innovation	22.975	32.277	35.587	35.062	46.054	42.007	29.296	45.061	40.635	46.006	29.960	35.981
Contribution of the intraday jumps												
J_y^+	3.083	1.659	4.48	3.921	3.281	3.202	6.086	3.268	2.232	2.354	2.736	3.762
J_y^-	2.958	2.588	3.518	2.896	3.769	1.865	3.688	2.299	1.752	2.1	2.137	2.756
J_m^+	0.641	0.481	1.334	0.951	1.371	2.081	0.924	2.221	1.711	2.795	0.485	1.333
J_m^-	0.662	0.297	1.244	0.982	1.503	1.428	1.513	1.077	1.992	0.976	0.415	1.196
Intraday jumps	7.344	5.025	10.576	8.750	9.924	8.576	12.211	8.865	7.687	8.225	5.773	9.047
Contribution of the overnight return												
Overnight return	7.992	3.698	5.960	5.931	5.098	8.435	4.005	9.152	15.357	3.806	5.836	9.762

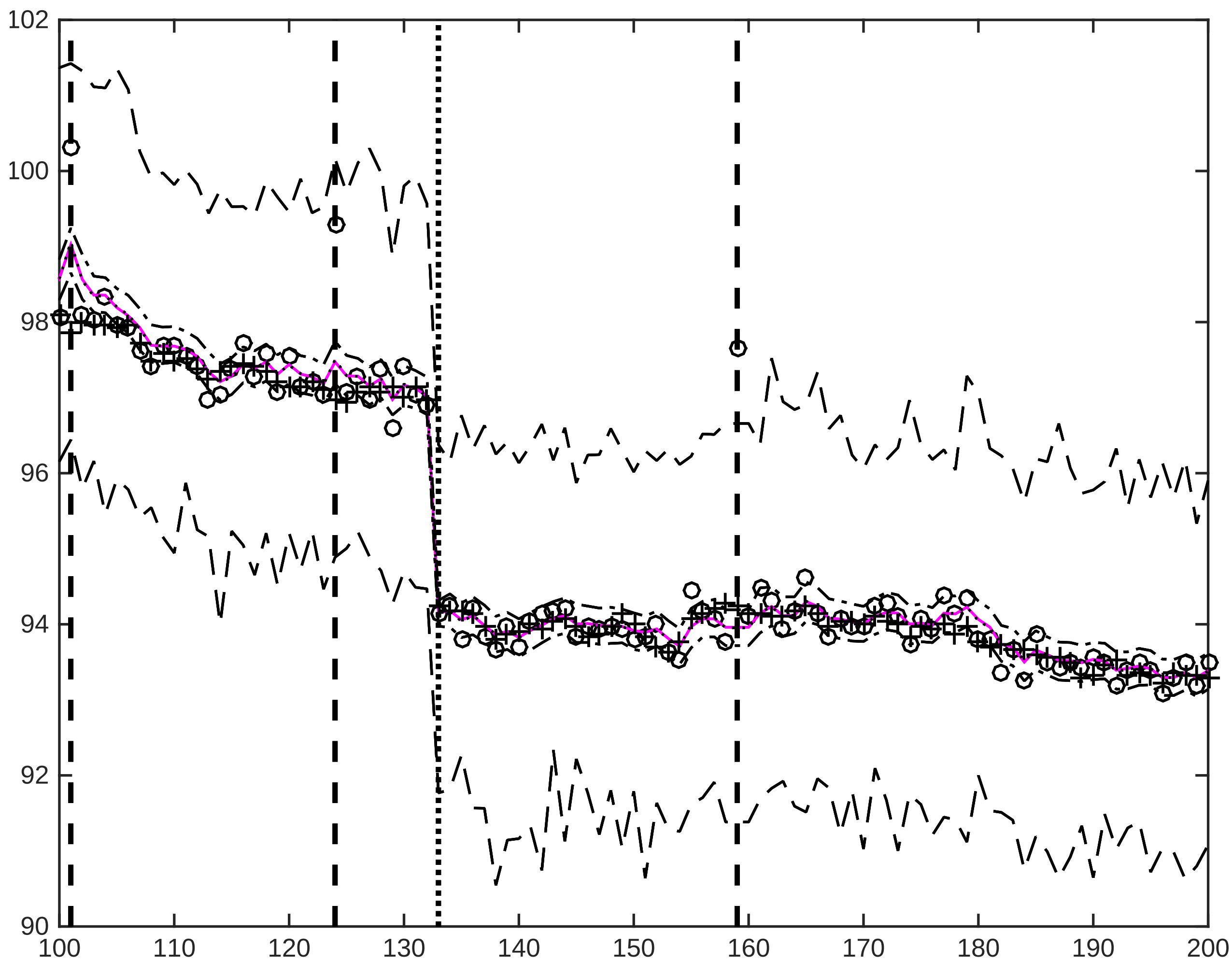
Note: The table reports the contribution of the various components of the model to the total variation in the stock return. See Section 4.3 for the construction of the variables.

Table 8: Price impact estimates

	Alcatel	Alstom	AXA	Fr. Tel.	Lagardere	LVMH	Orange	Sodexho	STMicro	Suez	Total	Vivendi
Panel A: Kalman filter												
Buy-initiated trade	6.121	12.250	5.099	4.665	6.032	4.617	6.438	7.410	2.746	5.313	3.920	5.375
Permanent impact	2.051	2.924	2.407	2.262	1.850	1.575	2.862	2.385	1.604	2.410	1.795	2.538
Transitory impact	4.070	9.326	2.692	2.403	4.182	3.043	3.576	5.025	1.142	2.904	2.124	2.837
Sell-initiated trade	-6.404	-16.654	-5.045	-3.898	-6.566	-4.233	-6.350	-7.701	-3.296	-5.374	-3.725	-5.030
Permanent impact	-1.751	-7.404	-2.487	-1.654	-1.716	-1.482	-1.842	-2.653	-1.888	-2.156	-1.573	-2.215
Transitory impact	-4.653	-9.249	-2.558	-2.243	-4.849	-2.751	-4.509	-5.048	-1.407	-3.218	-2.152	-2.816
Panel B: Particle Filter												
Buy-initiated trade	7.062	12.931	5.817	5.356	8.128	5.590	8.013	7.864	3.615	6.222	4.335	6.192
Permanent impact	2.489	3.995	3.028	2.769	3.491	2.291	3.943	3.652	2.237	3.330	2.268	3.339
Transitory impact	4.573	8.936	2.790	2.588	4.636	3.298	4.070	4.212	1.378	2.892	2.067	2.853
Sell-initiated trade	-7.210	-17.516	-5.747	-4.577	-7.132	-4.851	-6.818	-10.001	-4.017	-6.210	-4.058	-5.843
Permanent impact	-2.275	-7.677	-2.999	-2.080	-2.577	-2.166	-2.299	-3.994	-2.532	-3.009	-1.997	-2.826
Transitory impact	-4.935	-9.839	-2.748	-2.497	-4.556	-2.685	-4.519	-6.008	-1.484	-3.201	-2.061	-3.018

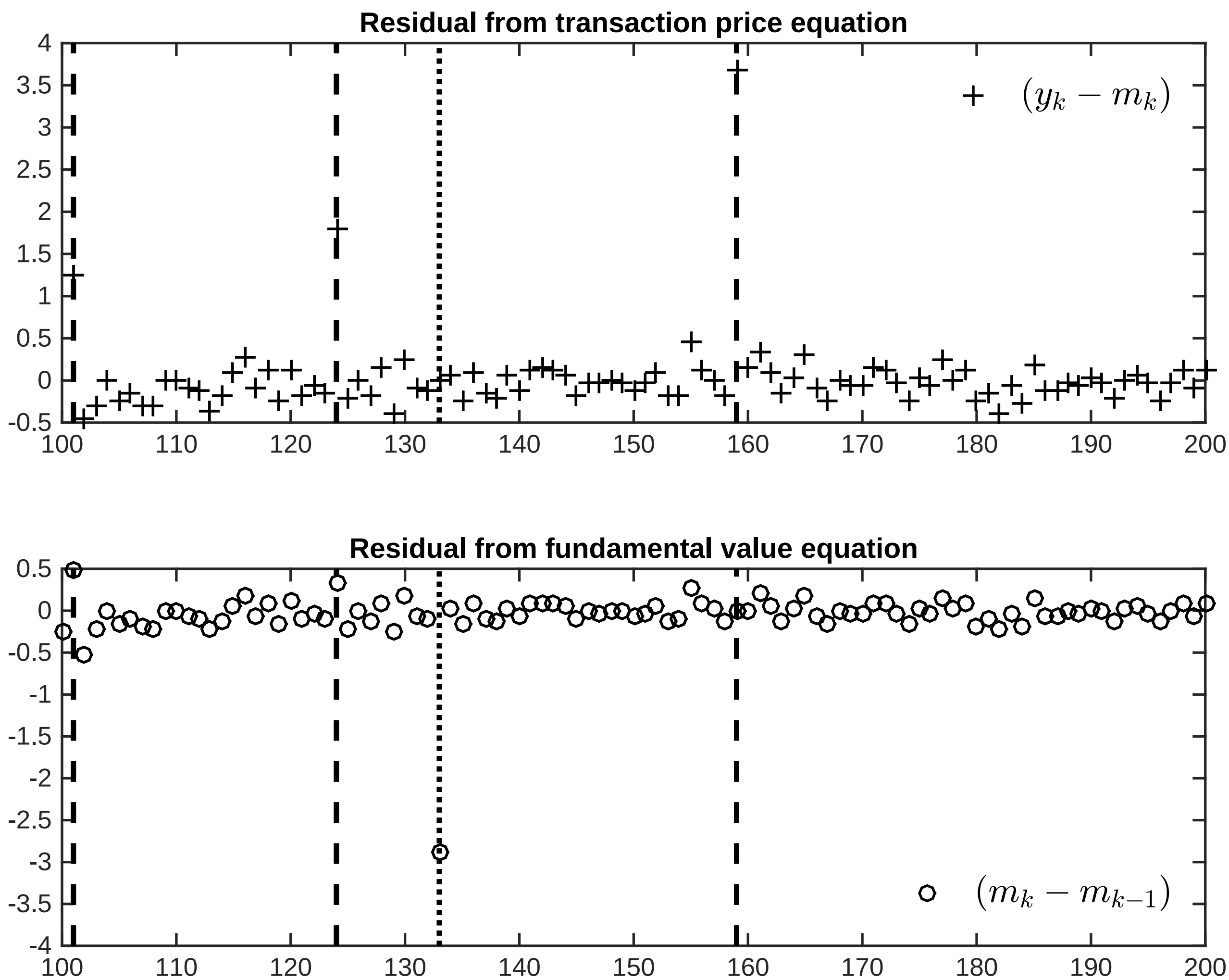
Note: The table reports price impacts (in basis points) resulting from the (buy-initiated or sell-initiated) trade of 10 times the average trade of the stock.

Figure 1: Price level y with jumps indicators



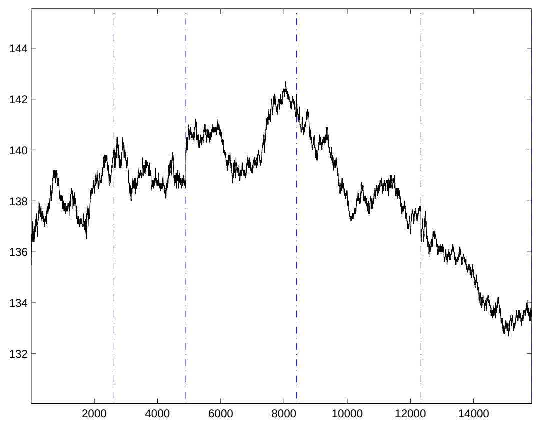
Note: The figure displays selected actual observations y_k obtained in a simulation exercise (o) as well as the corresponding true states (+). It also contains the 95% confidence interval concentrated around the particle-filter estimate of the state m_k (dash and dot) as well as of the posterior distribution of the observation, y_k (dashed line). Thin vertical lines indicate occurrence of permanent jumps (short dashes) or transitory jumps (long dashes). At observation 133, the algorithm detected an permanent jump in the state equation. At observations 101, 124, and 159, the simulated data contains a transitory outlier, which is also identified as such. The continuous line in the center corresponds to the median estimate of the state.

Figure 2: Residuals for observation and state equation and jump indicators

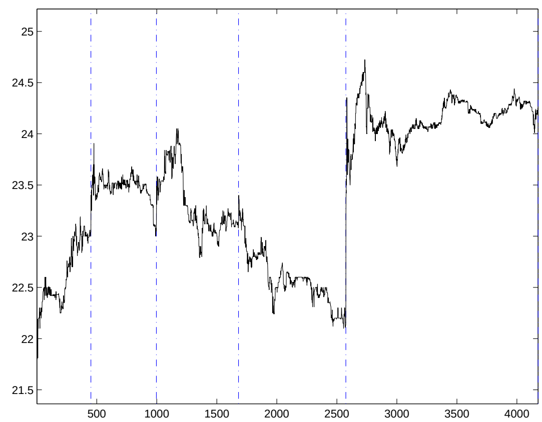


Note: The figure represents for simulated data the residuals of the observation and state equations, $(y_k - m_k)$ and $(m_k - m_{k-1})$, respectively. Inspection of the upper figure reveals that observations 101, 124, and 159 are large outliers. The large deviation of $(m_k - m_{k-1})$ for observation 133 leads to a successful detection of permanent jump.

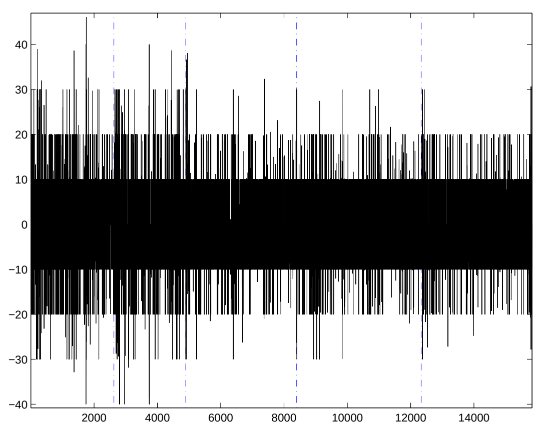
Figure 3: Price and price increment in tick time (5 days)



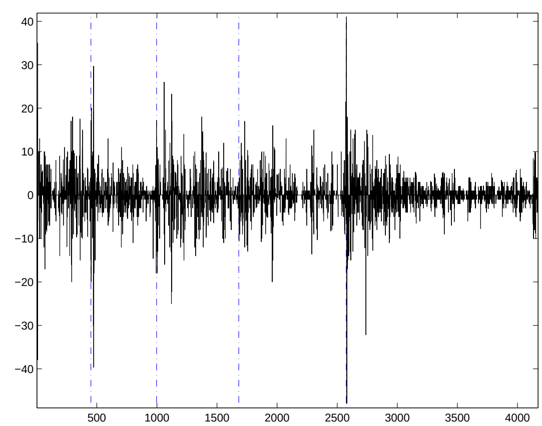
(a) Liquid stock (Total)



(b) Illiquid stock (Sodexho)



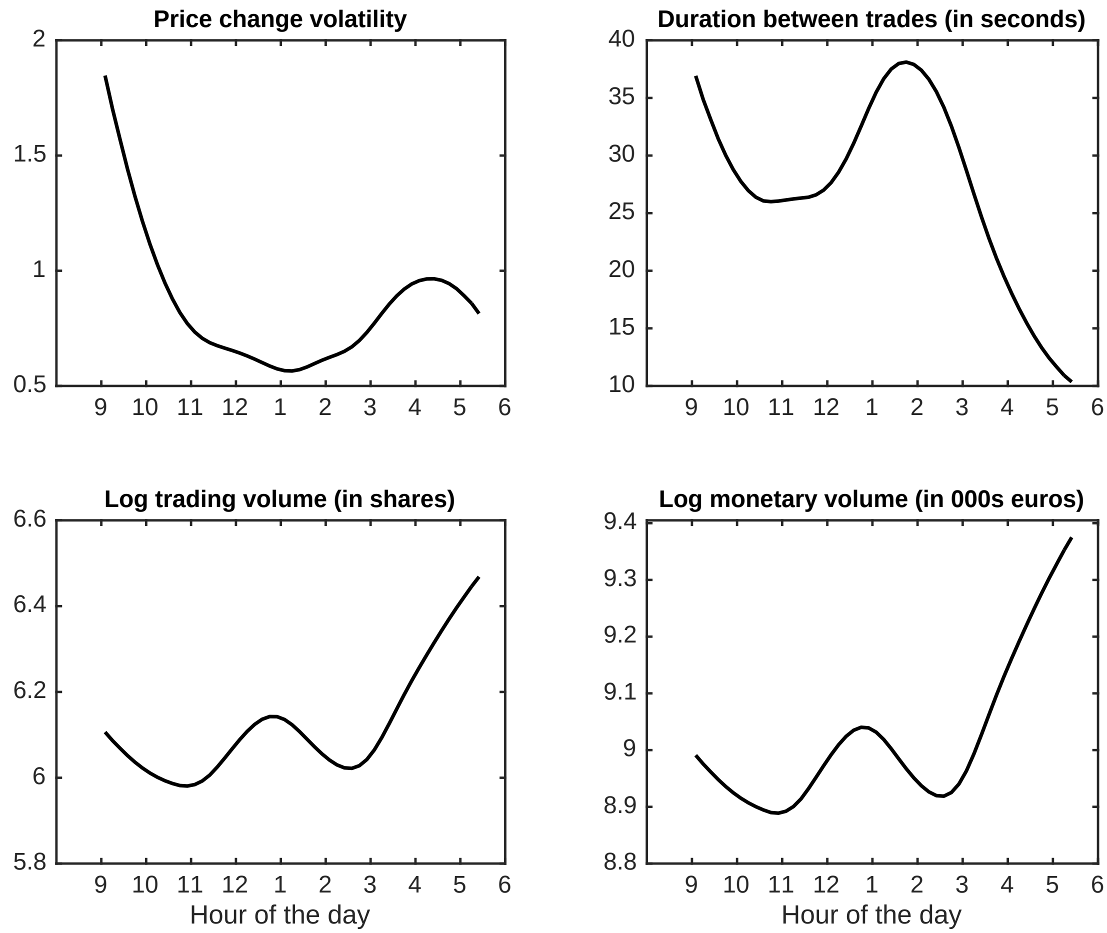
(c) Liquid stock (Total)



(d) Illiquid stock (Sodexho)

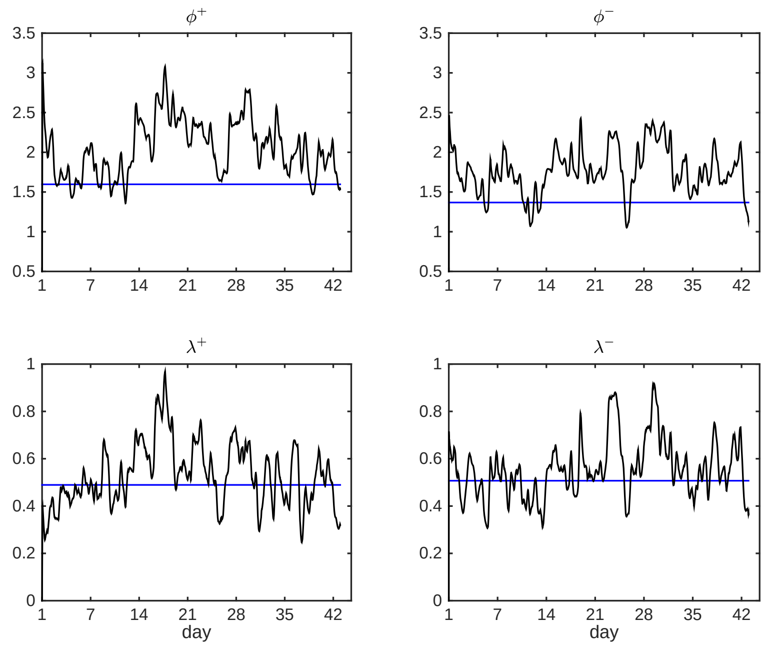
Note: The figure represents in (a) the price process (in euro) for a highly liquid company (Total) and in (b) the price process for a less liquid company (Sodexho). It also represents in (c) the price increment (in cents of euro) for Total and in (d) the price increment for Sodexho. The figures present data for 5 days (2, 3, 6, 7, and 8 January, 2003), each being separated form the next one by some vertical line.

Figure 4: Intraday seasonality.

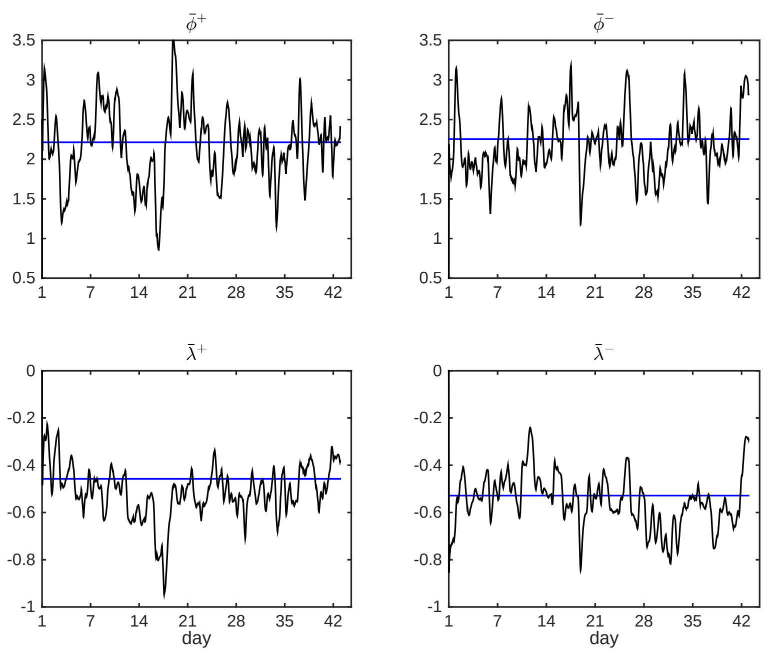


Note: The figure represents intradaily seasonality coefficients for the price change volatility, the duration between trades, and the log trading volume (in shares and in euro). For the volatility, the measure is obtained by using a jump-robust non-parametric estimates based on a multi-power volatility estimation. The figures correspond to the average across the stocks of the seasonality coefficients.

Figure 5: Parameter evolution of the complete model for Total.



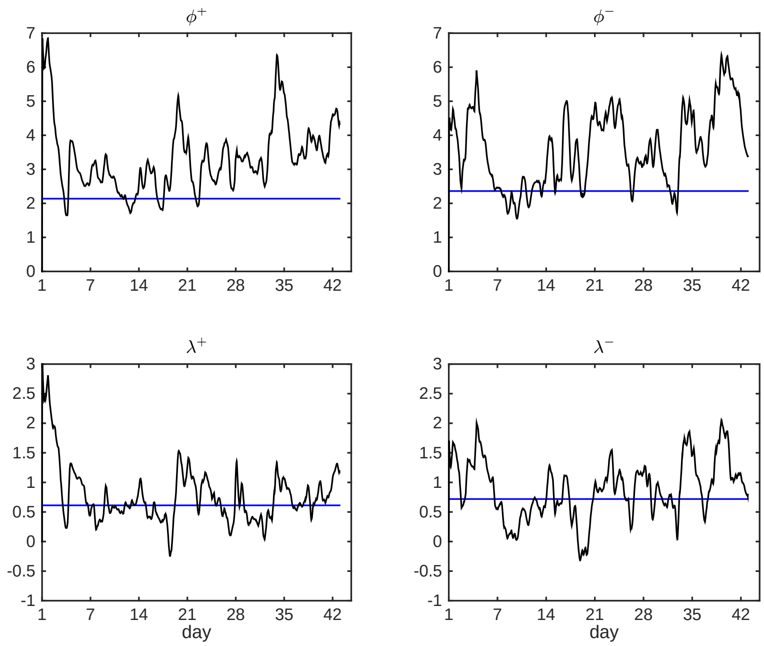
(a) Parameters of the fundamental value equation



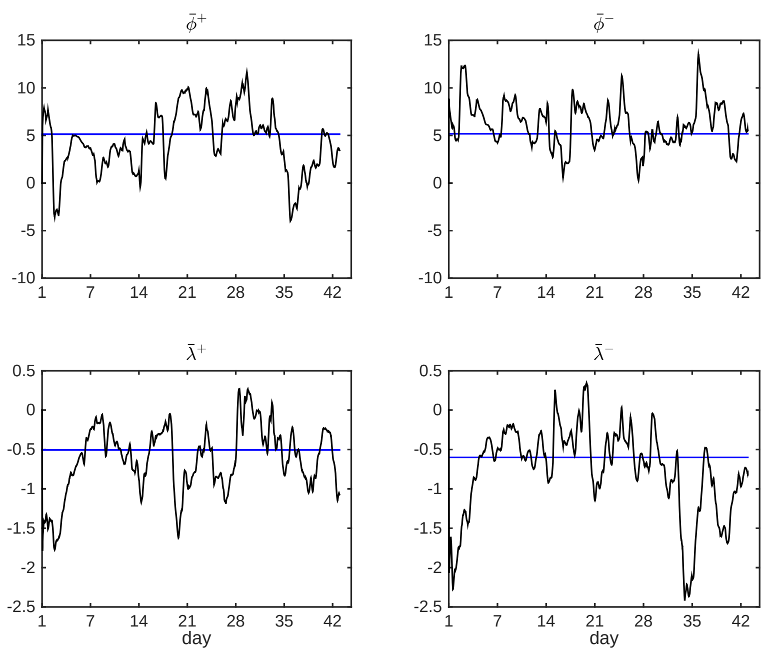
(b) Parameters of the transaction price equation

Note: The figure represents the parameter estimates obtained in an online estimation with daily re-initialization as described in the main text. The straight line corresponds to the Kalman-Filter estimates. Here, we represent the parameters contributing to the price impact of trades. The data is filtered for intradaily seasonality.

Figure 6: Parameter evolution of the complete model for Sodexho.



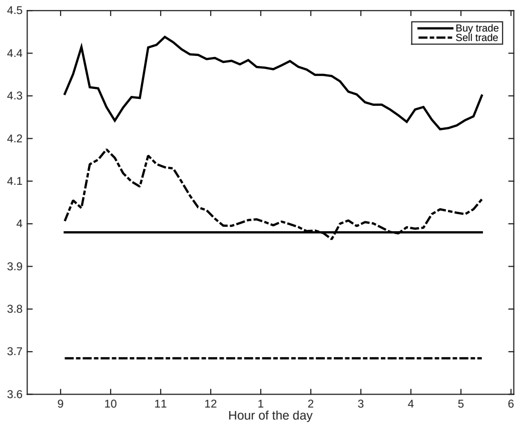
(a) Parameters of the fundamental value equation



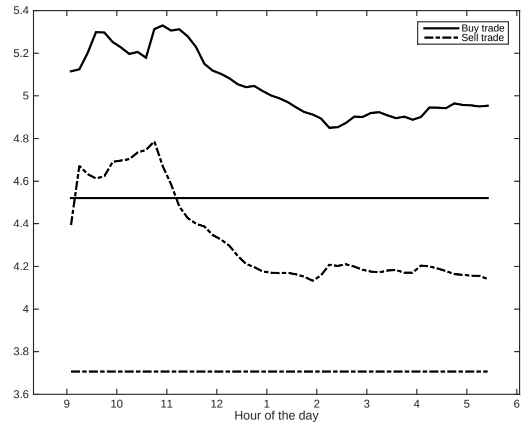
(b) Parameters of the transaction price equation

Note: The figure represents the parameter estimates obtained in an online estimation with daily re-initialization as described in the main text. The straight line corresponds to the Kalman-Filter estimates. Here, we represent the parameters contributing to the price impact of trades. The data is filtered for intradaily seasonality.

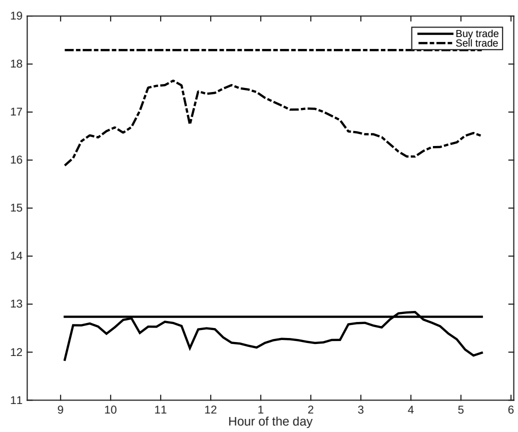
Figure 7: Price impact of trades for two highly liquid and two illiquid stocks.



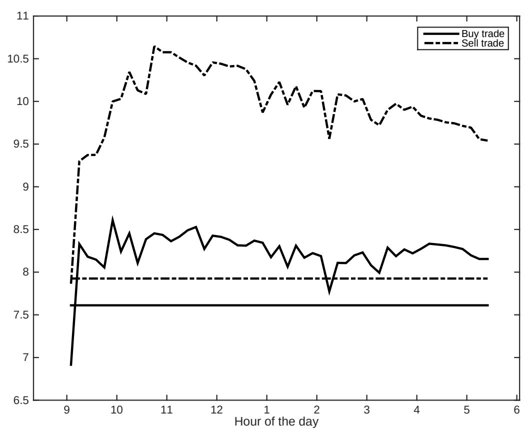
(a) Total



(b) France Telecom



(c) Alstom



(d) Sodexho

Note: The figure represents the price impact of a (buy-initiated or sell-initiated) trade of 10 times the average trade of the stock. The solid line corresponds to a buy trade, the dashed line to a sell trade. The straight line corresponds to the Kalman-Filter estimates.