Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN ECONOMIA

Ciclo XXIII

Settore Consorsuale di Afferenza: 13/A5 - ECONOMETRIA
Settore Scientifico Disciplinare: SECS-P05 - ECONOMETRIA

Topics in financial econometrics

Presentata da: Fabrizio Ferriani

Coordinatore Dottorato
Giacomo Calzolari

Relatore
Sergio Pastorello

Esame finale anno 2012

To my family
To Serena
To professor Sergio Pastorello, an incomparable master

Contents

Chapter 1

Introduction

This thesis consists of four self-contained chapters and its contribution to the literature splits between financial econometrics and the microstructure of financial markets. In the past years, the growing interest for the analysis of financial markets has fostered a considerable amount of theoretical and empirical research in these fields.

In the first chapter, we consider the joint estimation of objective and risk-neutral parameters for stochastic volatility option pricing models using both stock and option prices. This topic has been broadly investigated and it is particularly relevant for pricing and hedging of derivatives. A common strategy simplifies the task by limiting the analysis to some proxy of the latent volatility state or to a single option per date, by assuming that its price is observed without measurement error. This approach is particularly appealing as it allows to straightforwardly evaluate the loglikelihood with an application of the Jacobian formula. However, it presents some drawbacks and limitations, and we show that the estimates of the parameters highly depend on the choice of the option exempt from observation noise. Therefore, we propose an alternative strategy which exploits the wealth of information contained in large heterogeneous panels of options, and we apply it to S&P 500 index and index call options data. Our approach breaks the stochastic singularity between contemporaneous option prices by assuming that every observation is affected by measurement error. In this case, the evaluation of the likelihood function poses some non trivial numerical challenges, but we successfully overcome them by using a MC-IS strategy combined with a Particle Filter algorithm. The results we obtain confirm the validity of our method, though some significant improvements could be achieved by using a more flexible specification which allows jumps or regime switching in the volatility dynamics.

The second chapter examines the impact of different categories of traders on market transactions, which represents a subject of primary interest for the research devoted to informational issues in financial markets. We split market participants between informed traders, who are associated with institutional operators, and uninformed traders who embrace retail investors. We estimate a model which takes into account traders' identities at the transaction level, and we find that the stock prices follow the direction of institutional trading. Our results show that informed buyers exert a positive pressure to market prices when they trade with uninformed sellers, while the opposite holds for informed sellers trading with uniformed buyers. These results are particularly appealing as our empirical application is carried out with data from Euronext Paris which operates in a regime of anonymity. To explain our estimates, we examine the informativeness of a wide set of observed market variables and we find that most of them are highly and unambiguously significant to infer the identity of traders.

The third chapter provides an empirical contribution where particular emphasis is placed on the arrival time of market events. The analysis of financial durations has attracted considerable research attention and it represents an additional area of investigation for the study of informational issues. In this chapter, we investigate the relationship between the categories of market traders and three alternative definitions of financial durations. We consider trade durations, which are indicative of market activity, as well as price and volume durations which are well-suited for the testing of microstructure hypotheses. We adopt a Log-ACD model where we include information on traders at the transaction level, and we explore how informed traders and the liquidity provider affect the arrival of market events. As to trade durations, we observe an increase of the trading frequency when informed traders and the liquidity provider intensify their presence in the market. On the other hand, we find that the same effect for price and volume durations depends on the state of the market activity. Indeed, informed traders and the liquidity provider foster the arrival of the next (price or volume) spell, but only during periods of high trading frequency. These results provides an empirical confirmation of information models which theorize an accelerating effect for informed trading. Our estimates prove to be robust across alternative distributions, as well as when they are tested with supplementary microstructure variables.

Finally, in the fourth chapter we focus on orders aggressiveness at Euronext Paris. In the empirical microstructure literature, there exists a plenty of contributions which examine this topic in several financial markets worldwide. Aggressiveness is strictly related to the strategy of order submission and it is commonly evaluated through the classification introduced by Biais et al. (1995). A standard approach applies discrete response models to this ranking and it is particularly attractive for its simplicity, but also because it allows to deal with price discreteness. However, when the last issue is negligible, the use of categorical models is quite restrictive, as it collapses both price and volume informativeness. Therefore, we propose an alternative strategy where we replicate the classification of Biais et al. (1995), but we express order aggressiveness in quantitative terms. We consider a simultaneous equation model for price and volume aggressiveness at Euronext Paris that represents an interesting microstructure setting for the study of order aggressiveness. Indeed, our research is the first to investigate this topic in an order-driven market with liquidity providers and hidden identities. We examine a wide set of order book variables and we find evidence of autocorrelation patterns and intraday cycles for price and volume aggressiveness. Results show that price aggressiveness is mainly influenced by depth at best quotes, volatility, spread, and return, while volume aggressiveness is especially affected by volatility and spread. In the end, we also find the most aggressive orders to exert a higher impact on the stock prices.

Chapter 2

Estimating and Testing Non-Affine Option Pricing Models With a Large Unbalanced Panel of Options

Fabrizio Ferriani and Sergio Pastorello

JEL codes: C15, C58, G13

Keywords: Stochastic volatility, Jump-diffusion, Option pricing, Simulated Maximum Likelihood, Importance Sampling, Generalized residuals

2.1 Introduction

In this paper, we consider joint estimation of objective and risk-neutral parameters for non-affine jump-diffusion stochastic volatility (SV) option pricing models using both stock and option prices. This problem has been the subject of much work in recent empirical financial econometrics.

A common strategy simplifies the task by limiting the analysis to just one option per date, instead of the full cross-section, and assuming that its price is observed without measurement error. In this set up, there exists a one-to-one relationship between the observed variables and the state variables, and this makes the latter effectively observable. As a consequence, the loglikelihood can be evaluated using the Jacobian formula. The same result can be obtained using, instead of a single option price, some proxy of the latent volatility state, that can e.g. be derived using the VIX index as a proxy of the risk-neutral expectation of the integrated variance, and neglecting any noise it may contain. The simplicity of this approach explains its widespread adoption in the literature – see e.g. Aït-Sahalia and Kimmel (2007) and (2010) for an application to SV models and to term structure models, respectively.

It should be noted, however, that the assumption about which specific option is exempt from observation noise is essentially arbitrary, and that in principle many alternative and equally reasonable decisions would be possible. Some recent papers (e.g., Jiang and Tian, 2007), moreover, point to some systematic biases in the VIX. In either case, the estimates of the model parameters, and the filtered state variables and pricing errors, will in general depend on the assumptions made to recover the latent variables by inverting the model-implied expressions of the observable variables. This approach can also be problematic to implement in models in which the latent state variables are restricted to belong to a subset of the real line, a constraint that is not automatically satisfied by the inversion technique. Finally, it does not allow to price an option conditioning on more than just one option observed at the same date.

In this paper, we develop an alternative inference strategy that does not need assumptions of this kind. Such a procedure has already been advocated by Tauchen (2002) and Bates (2003). Its features can be summarized as follows. We break the stochastic singularity between contemporaneous option prices by assuming that every observation is affected by measurement error. We deem this assumption more appealing than the above one. The price to pay for this increased flexibility is that the evaluation of the likelihood function poses some non trivial numerical challenges, but we overcome them using a MC-IS strategy, combined with a Particle Filter algorithm along the lines suggested by Durham and Gallant (2002) and Durham (2007). We approximate the theoretical model-implied option prices using a highly flexible parametric model, which allows us to compute quickly and accurately a huge number of prices.

For readability, the discussion is based on a set of simplifying assumptions, but it is important to remark that many of these could be relaxed without difficulties, as they are not essential for the implementation of our approach. In particular, different assumptions concerning risk premia structures or jump intensities and size distributions could be handled fairly easily, and ML inference would still be feasible. Notice, however, that a few existing contributions already considered some of these extensions with mixed results. For example, Bates (2000) considers a specification in which the jump intensity can be an affine function of the volatility state, but fails to reject the null of a zero slope. A constant jump intensities is also supported by Chernov et al. (2003) and Andersen et al. (2001).

The paper is structured as follows. The following section outlines the option pricing model, and provides details on the specification that we actually consider in the empirical analysis. It also discusses some problems that are commonly encountered in the literature of empirical option pricing, such as the need to approximate the transition density of the state variables and the model-implied theoretical option pricing formula. Section 2.3 discusses the simpler strategy based on the assumption that one option at each date is devoid of measurement error, and on the Jacobian formula to compute the loglikelihood. We highlight its drawbacks and provide an empirical illustration. Section 2.4 outlines the alternative approach which assumes measurement errors on each option. We first describe the strategy we use to approximate the loglikelihood, and show that with some minor modifications of the same techniques we can also easily compute filtered values of the state variables and of functions thereof that can be extremely useful in testing the specification of the model and in using it for pricing purposes. Section 2.5 illustrates an application of this approach to a sample of call options on the S&P 500 equity index. Finally, section 2.6 concludes. Details about the sample and the methodology employed in the paper are provided in the appendices.

2.2 Option pricing under jump-diffusion stochastic volatility

2.2.1 The model

In a jump-diffusion SV model, the dynamics under the risk-neutral probability measure $\mathbb{Q}$ of the price S_t of the underlying asset and the associated volatility state V_t is described by:

$$\frac{dS_t}{S_t} = (r_t - d_t)\,dt + \sigma_S(V_t)\,dW_{St}^{\mathbb{Q}} + (e^{J_t^{\mathbb{Q}}} - 1)\,dN_t^{\mathbb{Q}} - \lambda_J^{\mathbb{Q}}(e^{\nu_J^{\mathbb{Q}}} - 1)\,dt \tag{2.1}$$

$$dV_t = \mu_V^{\mathbb{Q}}(V_t)dt + \sigma_V(V_t)dW_{Vt}^{\mathbb{Q}} \tag{2.2}$$

where r_t and d_t are the instantaneous riskless interest rate and dividend rate, respectively, and, under $\mathbb{Q}$, $W_{St}^{\mathbb{Q}}$ and $W_{Vt}^{\mathbb{Q}}$ are standard Brownian motions with instantaneous correlation ρ, $N_t^{\mathbb{Q}}$ is a Poisson process with intensity $\lambda_J^{\mathbb{Q}}$, and $J_t^{\mathbb{Q}}$ is the jump size. We assume that the jump intensity and the jump size are independent from each other and from every other variable in the model, and that $J_t^{\mathbb{Q}} \sim \mathcal{N}(\mu_J^{\mathbb{Q}}, {\sigma_J^{\mathbb{Q}}}^2)$. Finally, we denote $\nu_J^{\mathbb{Q}} = \mu_J^{\mathbb{Q}} + {\sigma_J^{\mathbb{Q}}}^2/2$.

To derive the dynamics of the state variables under the objective measure $\mathbb{P}$, we need some assumptions about the structure of the risk premia. In this paper, we assume that the return risk premium on the Brownian shocks is given by $\eta\sigma_S(V_t)$, where $\sigma_S(V_t)$ is the diffusion coefficient in the price process – see (2.3) below – and η is a constant parameter. The volatility and jumps-related risk premia could also be specified explicitly; however, following Broadie et al. (2007), we simply specify a different dynamics of V, as well as different jump intensity and jump size distribution, and we interpret the difference between the $\mathbb{P}$ and $\mathbb{Q}$ parameters as risk premia.

We do not impose a priori any constraint on the volatility risk premium; the specific functional forms of $\mu_V^{\mathbb{Q}}(V_t)$ and $\mu_V(V_t)$ adopted in the empirical applications below allow to keep some simplicity in the model and are coherent with previous work in the field.

Under the previous assumptions, the dynamics of the state variables under $\mathbb{P}$ is given by:

$$\begin{aligned}\frac{dS_t}{S_t} &= [(r_t - d_t) + \eta\sigma_S(V_t)^2 - \lambda_J^{\mathbb{Q}}(e^{\nu_J^{\mathbb{Q}}} - 1)]dt + \sigma_S(V_t)dW_{St} + (e^{J_t} - 1)\, dN_t, &(2.3)\\ dV_t &= \mu_V(V_t)dt + \sigma_V(V_t)dW_{Vt}, &(2.4)\end{aligned}$$

where, under $\mathbb{P}$, W_{St} and W_{Vt} are standard Brownian motions with instantaneous correlation ρ, N_t is a Poisson process with intensity λ_J, and $J_t \sim \mathcal{N}(\mu_J, \sigma_J^2)$. For estimation purposes, it is convenient to work with the log price P_t, whose dynamics can be easily derived from (2.3):

$$dP_t = [(r_t - d_t) + (\eta - 1/2)\sigma_P(V_t)^2 - \lambda_J^{\mathbb{Q}}(e^{\nu_J^{\mathbb{Q}}} - 1)]dt + \sigma_P(V_t)dW_{St} + J_t\, dN_t, \qquad (2.5)$$

notice that $\sigma_P(V_t) = \sigma_S(V_t)$. For further reference, let us denote with $\mu_P(V_t) = (r_t - d_t) + (\eta - 1/2)\sigma_P(V_t)^2 - \lambda_J^{\mathbb{Q}}(e^{\nu_J^{\mathbb{Q}}} - 1)$ the drift coefficient in (2.5).

Previous work in this area focussed mainly on models in the affine class due to tractability considerations and to the existence of quasi-closed form expressions for option prices. Several works, however, emphasized the conclusion that affine models can be frequently badly misspecified; see, among others, Christoffersen et al. (2010). In this paper, we consider non-affine models that seem to provide a better fit to the data, either thanks to a different assumption on the volatility process (the log volatility model), or to increased flexibility (the CEV model). More precisely, the models we consider can be obtained from the general specification above if we impose the following constraints on the unspecified drift and diffusion coefficients:

- Log volatility (LOG-J) model:

$$\sigma_S(V_t) = \exp(V_t/2), \quad \mu_V(V_t) = \alpha + \beta V_t, \quad {\mu_V}^{\mathbb{Q}}(V_t) = \alpha^{\mathbb{Q}} + \beta^{\mathbb{Q}} V_t, \quad \sigma_V(V_t) = \gamma$$

- Constant elasticity of variance (CEV-J) model:

$$\sigma_S(V_t) = \sqrt{V_t}, \quad \mu_V(V_t) = \alpha + \beta V_t, \quad \mu_V^{\mathbb{Q}}(V_t) = \alpha^{\mathbb{Q}} + \beta^{\mathbb{Q}} V_t, \quad \sigma_V(V_t) = \gamma V_t^{\varphi}$$

In both models the volatility drift is linear under either $\mathbb{P}$ and $\mathbb{Q}$. Many studies based on this model assumed a single free parameter in the volatility risk premium, which implies that both drift parameters change between $\mathbb{P}$ and $\mathbb{Q}$, but not independently. On the contrary, we adopt a more flexible specification with two free parameters in the volatility risk premium. This allows α and β to vary independently across probability measures.

The CEV-J model collapses to an affine specification under the constraint $\varphi = 1/2$. Affine models have attracted a huge amount of attention in the literature, and we also considered this specification in the analysis. Given that this specification is overwhelmingly rejected by the data, and to save space, we do not report the corresponding results, and we limit our discussion to the LOG-J and CEV-J models. Finally, it should be noted that the LOG-J and CEV-J specifications can not be embedded into the affine class through the use of an augmented state, as it is the case, for example, for the Linear Quadratic Jump Diffusion models examined in Cheng and Scaillet (2007).

These specifications are more general than those that have been considered in the literature so far; for example, Broadie et al. (2007) fit on a large cross-section of S&P futures option prices from 1987 to 2003 an affine model in which the jump risk premia are similar to ours, but with a constrained risk premium specification on the volatility process; Durham (2010) considers the same non-affine models we do, but constrains the jump related parameters to be the same across the two probability measures. In the empirical implementations we will also sometimes consider constrained specifications in which some of the jump parameters coincide under the two measures. The analysis in Durham (2010) is also based on the time series of S&P 500 index returns and the VIX index, and neglects the cross-sectional dimension in options data.

For the non-affine processes we consider the transition density $f(P_t, V_t|P_{t-1}, V_{t-1})$ is unknown, and must be approximated. Several strategies have been advanced to solve this problem, the most successful two being the closed-form Hermite polynomials expansion of Aït-Sahalia (2008) or the IS strategy developed by Durham and Gallant (2002). This paper is based on the latter because of its greater flexibility, which will be particularly convenient in the approach illustrated in section 2.4. Details on the IS approach we use are provided in Appendix 2.B.

2.2.2 Numerical evaluation of theoretical option prices

Since volatility is not observable, we use option prices to extract information on the latent state. Our sample is a highly unbalanced panel of prices of European call options that for each observation date differ by strike price and/or time to maturity. Following Bates (2000), we focus on option prices normalized by the underlying asset price discounted at the dividend rate. Consider the generic i-th option observed at date t, and let C_{it}, K_{it} and τ_{it} denote its price, strike and time to maturity, respectively. Moreover, let r_t and d_t be the instantaneous riskless interest rate and dividend rate. The option's normalized price (NP) is then defined as:

$$H_{it} = \frac{C_{it}}{S_t e^{-d_t \tau_{it}}}.$$

We collect in $\mathbf{H}_t = (H_{it},\, i = 1, \ldots, N_t)$ the $(N_t \times 1)$ NPs at date t. Define $X_{it} = \ln(S_t/K_{it}) + (r_t - d_t)\tau_{it}$ as the log discounted moneyness of the option, and let $\boldsymbol{\chi}_{it} = (X_{it}, \tau_{it})'$ be the vector of the option's characteristics. We denote with $h(V_t, \boldsymbol{\chi}_{it})$ the model implied theoretical NP. Notice that to simplify the notation we omit the occurrence of $\boldsymbol{\theta}$ in h.

Appendix 2.C illustrates a numerical technique that can be used to evaluate option prices in non-affine jump diffusion SV models. Even if relatively fast, this approach is still too slow for our sample size. For this reason we approximate $h(V_t, \boldsymbol{\chi}_{it})$ using a polynomial interpolation scheme. We first construct a fixed three-dimensional grid

$\{(V_g, \boldsymbol{\chi}_g')' = (V_g, X_g, \tau_g)', \, g = 1, \ldots, G\}$ combining three univariate grids, spanning the range of variation of the corresponding variable. Given a value for $\boldsymbol{\theta}$, we evaluate the theoretical option price at each point on the grid, H_g, using the approach described in Appendix 2.C. We then use the set $\{(H_g, V_g, X_g, \tau_g)', \, g = 1, \ldots, G\}$ to construct an interpolation scheme that approximates $\log H_g$ with a polynomial in $(\log V_g, X_g, \tau_g)'$ for several reasons. First, its coefficients can be computed very quickly and accurately by OLS. Second, given the estimates, it is immediate to compute an approximation of $h(V_t, \boldsymbol{\chi}_{it})$ and of its derivative with respect to V_t, which is needed in the empirical applications below. We use a polynomial of order four with 35 parameters. To estimate them, we consider equally spaced univariate grids with 10 points for $\log V$, and 6 points for τ and X. The three-dimensional grid contains $G = 360$ points. The R^2 coefficient of the interpolating regression is always larger than 0.998.

A natural alternative to numerical schemes would be to use one of the recently advanced analytical approximations of the theoretical NPs in jump diffusion SV models (see e.g. Lewis, 2000, Sircar and Papanicolau, 1999, Lee, 2001, and Medvedev and Scaillet, 2007), which provide extremely fast tools to evaluate theoretical NPs. In this paper, however, we prefer the latter because some preliminary Monte Carlo experiments highlighted that the quality of approximation characterizing the analytical expansions is lower than that of the numerical scheme.

2.2.3 Measurement errors

For any candidate $\mathbb{Q}$, the pricing model states that the NP of any option is a function of $(S_t, V_t)'$, which in turn implies the existence of a set of exact relations between the NPs of different options at the same date. This conclusion is rejected in any data set. To overcome this issue we could consider just one option per date, but this amounts to neglect a huge amount of information on the latent state. Moreover, the choice of the specific single option to be considered at each date would be, to a large extent, essentially arbitrary.

Alternatively, the stochastic singularity can be broken by introducing additional sources of statistical uncertainty. Increasing the dimension of the state vector would be theoretically sound but extremely complicated. For this reason, the solution usually adopted is to assume that option prices are observed with an error that can be due to microstructure effects (e.g., bid-ask spreads and tick-by-tick price variations) and data issues (e.g., non synchronous or only approximate observation of the relevant variables). Measurement errors in option prices can be assumed implicitly, e.g. when parameters are estimated through least squares techniques, or explicitly, as a component of the estimation strategy.

ML inference requires an assumption about the stochastic structure of the observation errors. In this paper, we assume additive measurement errors in log NPs, defined by:

$$\varepsilon_{it} = \log H_{it} - \log h(V_t, \boldsymbol{\chi}_{it}),$$

distributed independently through time and across options according to a Gaussian distribution with mean zero. We also allow for some heteroskedasticity by maintaining that:

$$\omega_{it}^2 = \mathrm{Var}(\varepsilon_{it}|V_t, \boldsymbol{\chi}_{it}) = \exp[\psi_0 + \psi_X X_{it} + \psi_{X^2} X_{it}^2 + \psi_\tau(\tau_{it}/365) + \psi_{\tau^2}(\tau_{it}/365)^2].$$

We merge in $\boldsymbol{\theta}$ the parameters appearing in the measurement error distribution. The assumption of independence across dates and options is not essential, but we think that it is reasonable: (i) it limits the number of nuisance parameters, and (ii) we believe that any correlation between options should be accounted for by the pricing model, and not by the measurement errors. Our assumptions are also largely confirmed by the empirical results below. Finally, the techniques we analyze could be extended to handle different definitions of the errors or of their distribution, including alternative forms of heteroskedasticity depending on V_t.

2.3 Volatility filtering by option prices inversion

2.3.1 Loglikelihood derivation using the Jacobian formula

A common approach assumes that at each date exactly one option is observed without error, whereas the remaining $N_t - 1$ are affected by measurement noise. This choice equalizes the dimensions of the augmented vector of latent variables (volatility and measurement errors) and of the observed option prices, allowing to derive the likelihood contribution with an application of the Jacobian formula.

To illustrate this strategy, let us partition the observed NPs as in $\mathbf{H}_t = (H_{1t}, \mathbf{H}'_{2t})'$, where, without loss of generality, H_{1t} is assumed to be noise free, whereas $\mathbf{H}_{2t}$ are $N_t - 1$ NPs affected by error. Furthermore, let us denote with $f_{PV}(P_t, V_t | P_{t-1}, V_{t-1})$ the transition density of the log price and its volatility derived from (2.5)-(2.4). The Jacobian formula then states that the transition density of the observables is given by:

$$f(P_t, H_{1t} | P_{t-1}, H_{1t-1}) = f_{PV}[P_t, h^{-1}(H_{1t}, \boldsymbol{\chi}_{1t}) | P_{t-1}, h^{-1}(H_{1t-1}, \boldsymbol{\chi}_{1t-1})] \times \left| \frac{\partial h^{-1}(H_{1t}, \boldsymbol{\chi}_{1t})}{\partial H_{1t}} \right|$$

which is the date t likelihood contribution of P_t and H_{1t}. The exact expression of the transition pdf is generally unknown, and in this paper we approximate it using the IS approach in Appendix 2.B. To derive the likelihood contribution of the remaining NPs, we observe that the measurement error on the options 2 to N_t is given by:

$$\varepsilon_{it} = \log H_{it} - \log h[h^{-1}(H_{1t}, \boldsymbol{\chi}_{1t}), \boldsymbol{\chi}_{it}].$$

Given our assumption of independent $\mathcal{N}(0, \omega_{it}^2)$ measurement errors, the density of $\mathbf{H}_{2t}$ conditional on V_t, or equivalently on H_{1t}, is given by $\prod_{i=2}^{N_t} \phi(\varepsilon_{it}; 0, \omega_{it}^2)$. The sample loglikelihood is then given by

$$\ell_T(\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{t=1}^{T} \left[\log f(P_t, H_{1t} | P_{t-1}, H_{1t-1}) + \sum_{i=2}^{N_t} \log \phi(\varepsilon_{it}; 0, \omega_{it}^2) \right]. \qquad (2.6)$$

2.3.2 Issues

The simplicity of this approach explains its widespread adoption – see e.g. Aït-Sahalia and Kimmel (2007) and (2010) for an application to SV models and to term structure models. Notice, however, that the choice of the option exempt from observation noise is arbitrary. In principle, many alternative and equally reasonable decisions would be possible. Moreover, the estimates of the parameters will in general depend on this choice; see the next section for an illustration in option pricing models. Finally, this approach can be problematic to implement in models in which the latent state variables are restricted to belong to a subset of the real line.

To elaborate on the latter point, notice that in a SV model the normalized price of an option can not be smaller than some lower bound. In the Black and Scholes model this lower bound is $\max[0, 1-\exp(-X)]$, and it is attained for a zero diffusion coefficient. When volatility is stochastic the bound is still attained for a zero V_t, but it is higher, and it depends on $\boldsymbol{\theta}$. For an option with characteristics $\boldsymbol{\chi}_{it}$, we denote the lower bound for the NP with $H_{LB}(\boldsymbol{\chi}_{it}, \boldsymbol{\theta})$. The crucial step in the previous approach, which effectively makes V observable, is to compute the solution of the $T+1$ nonlinear equations

$$H_{1t} = h(V_t, \boldsymbol{\chi}_{1t}), \quad \text{for } t = 0, 1, \ldots, T.$$

However, for these equations to admit a solution it is necessary that the following nonlinear inequality constraints be satisfied:

$$H_{1t} \geq H_{LB}(\boldsymbol{\chi}_{1t}, \boldsymbol{\theta}), \quad \text{for } t = 0, 1, \ldots, T. \tag{2.7}$$

Hence the ML estimation problem must be formulated as an optimization under $T+1$ nonlinear inequality constraints:

$$(\widehat{\boldsymbol{\theta}}', \widehat{\boldsymbol{\psi}}')' = \arg \max_{(\boldsymbol{\theta}', \boldsymbol{\psi}')' : \{H_{1t} \geq H_{LB}(\boldsymbol{\chi}_{1t}, \boldsymbol{\theta}),\, t=0,1,\ldots,T\}} \ell_T(\boldsymbol{\theta}, \boldsymbol{\psi})$$

The presence of a $T+1$ nonlinear constraints on the parameters greatly complicates inference; see Duffee (2002) for a discussion in affine term structure models. In practice, it is impossible to solve the estimation problem using standard techniques of maximization under constraints. The only feasible strategy consists in imposing a huge penalty to the loglikelihood whenever $\boldsymbol{\theta}$ does not satisfy some of the inequality constraints. In turn, this introduces large discontinuities in the objective function, which essentially prevent the use of derivative-based optimization algorithms. Even algorithms that do not require derivatives (such as the Simplex method we use) almost always get stuck on the boundary of the parameter space generated by one of the constraints; as a consequence, the end result is usually a boundary local maximum.

We now provide an empirical illustration of the above discussion. For simplicity, we focus on the effect of the choice of the option observed without error, and we neglect the numerical issues posed by the presence of the huge number of nonlinear constraints (2.7).

2.3.3 An application to S&P 500 options

In this section, we apply the NP inversion approach to a sample of options on the S&P500 stock index. We defer to appendix 2.A for a description of the data set. As option pricing models, we consider the LOG-J and the CEV-J jump diffusion SV models discussed in section 2.2.1.

Our purpose is to illustrate the impact on the parameters estimates and derived quantities of 5 possible criteria used to identify the noise free contract. For each criterion, we first select the options with time to maturity closest to some target value τ_*, and then pick among them the one with discounted moneyness closest to some value X_*. Different criteria correspond to different choices about the target τ_* and X_*.

- Criterion 1: $\tau_* = 30$, $X_* = 0$
- Criterion 2: $\tau_* = 15$, $X_* = 0$
- Criterion 3: $\tau_* = 60$, $X_* = 0$

- Criterion 4: $\tau_* = 30$, $X_* = -1\%$
- Criterion 5: $\tau_* = 30$, $X_* = +1\%$

The first criterion is a slight modification of the selection rule used by Pan (2002); its targets τ_* and X_* identify at-the-money short-lived options. Criteria 2 and 3 allow to appreciate the effect of a different choice of τ_*, while criteria 4 and 5 set the target time to maturity at 30 days, and consider slightly out-of-the-money and in-the-money options, respectively. Notice that the subsamples of options used to recover the volatility states by the 5 criteria partially overlap; in particular, criterion 2 selects just 99 (3.9%) different contracts with respect to criterion 1. The other criteria overlap to a lesser degree: the numbers (percentages) of different options for criteria 3, 4 and 5 are 1,309 (52%), 2,321 (92.2%) and 2,286 (90.8%), respectively.

Tables 2.1 and 2.2 illustrate the parameter estimates for the LOG-J and CEV-J models. The tables contain five columns, one for each of the criteria used to identify the NPs to be inverted. For each parameter, we report the SML estimates and the corresponding asymptotic standard error (in parenthesis) derived from the outer product of gradients estimate of the asymptotic variance matrix. To save space, we omit the estimates of the heteroskedasticity $\boldsymbol{\psi}$ parameters, but they are available at request. The parameter estimates are generally quite accurate, and in line with results reported elsewhere. The R^2 coefficients of the polynomial interpolation used to approximate option prices are always larger than 0.998.

Given the amount of information contained in our sample of options, it is not surprising that the drift and jump parameters under $\mathbb{Q}$ are estimated with much higher precision than the corresponding parameters under $\mathbb{P}$. For the CEV-J model, the estimates of φ are always significantly larger than 1, which is coherent with the results in Jones (2003), although our estimates are somewhat lower. As shown by Conley et al. (1997), a value of φ higher than 1 implies that the stationarity of the V_t process is "volatility-induced" irrespectively of the sign of β, provided that α is positive. The latter condition is always

Table 2.1: Estimates of the parameters for the LOG-J option pricing model on the sample of options on the S&P 500 index observed on each day from Jan. 4, 1996 to Dec. 30, 2005. Each column contains the estimates obtained by selecting the NPs to be inverted following one of the five criteria detailed in section 2.3.3. Asymptotic standard errors in parentheses.

	$\tau_* = 30$ $X_* = 0\%$	$\tau_* = 15$ $X_* = 0\%$	$\tau_* = 60$ $X_* = 0\%$	$\tau_* = 30$ $X_* = -1\%$	$\tau_* = 30$ $X_* = 1\%$
λ	3.67	3.01	7.86	-0.90	11.32
	(2.55)	(2.61)	(2.71)	(2.24)	(3.47)
α	-2.11×10^{-2}	-2.88×10^{-2}	-1.03×10^{-2}	-1.63×10^{-2}	-8.95×10^{-2}
	(7.25×10^{-3})	(9.21×10^{-3})	(4.48×10^{-3})	(6.12×10^{-3})	(1.75×10^{-2})
β	-4.10×10^{-2}	-4.66×10^{-2}	-2.85×10^{-2}	-2.68×10^{-2}	-6.57×10^{-2}
	(8.24×10^{-3})	(8.63×10^{-3})	(5.75×10^{-3})	(7.81×10^{-3})	(9.42×10^{-3})
γ	0.28	0.33	0.20	0.25	0.48
	(1.90×10^{-3})	(2.35×10^{-3})	(1.04×10^{-3})	(1.67×10^{-3})	(2.99×10^{-3})
ρ	-0.78	-0.75	-0.88	-0.75	-0.84
	(3.99×10^{-3})	(3.78×10^{-3})	(3.73×10^{-3})	(4.12×10^{-3})	(2.75×10^{-3})
$\alpha^{\mathbb{Q}}$	-8.56×10^{-3}	-1.71×10^{-2}	3.51×10^{-3}	-1.00×10^{-2}	6.77×10^{-3}
	(2.83×10^{-4})	(4.01×10^{-4})	(9.61×10^{-5})	(2.56×10^{-4})	(7.08×10^{-4})
$\beta^{\mathbb{Q}}$	-1.29×10^{-2}	-1.51×10^{-2}	-1.64×10^{-2}	-8.27×10^{-3}	-3.20×10^{-2}
	(4.35×10^{-5})	(5.90×10^{-5})	(4.31×10^{-5})	(5.12×10^{-5})	(6.48×10^{-5})
$\lambda_J^{\mathbb{Q}}$	3.51×10^{-2}	4.03×10^{-2}	2.82×10^{-2}	1.91×10^{-2}	0.24
	(4.19×10^{-4})	(4.41×10^{-4})	(3.41×10^{-4})	(5.50×10^{-5})	(8.87×10^{-4})
$\mu_J^{\mathbb{Q}}$	-0.54	-0.54	-0.32	-1.52	0.14
	(7.03×10^{-3})	(5.52×10^{-3})	(6.58×10^{-3})	(4.49×10^{-3})	(1.71×10^{-3})
λ_J	0.39	0.36	0.93	0.18	1.85
	(2.43×10^{-2})	(2.22×10^{-2})	(5.07×10^{-2})	(1.43×10^{-2})	(9.92×10^{-2})
μ_J	-8.81×10^{-2}	-8.38×10^{-2}	-0.10	2.07×10^{-2}	-9.14×10^{-3}
	(0.12)	(0.15)	(5.33×10^{-2})	(0.35)	(3.48×10^{-2})
σ_J	1.81	1.98	1.18	2.88	1.02
	(2.03×10^{-2})	(1.94×10^{-2})	(1.18×10^{-2})	(5.56×10^{-3})	(2.34×10^{-3})
loglik.	32040.3	31599.6	30839.5	32428.9	29473.5

satisfied under $\mathbb{P}$, but violated under $\mathbb{Q}$. Hence, according to the estimates in table 2.2, the volatility process is nonstationary under the risk-neutral measure.

Inspection of tables 2.1 and 2.2 highlights several discrepancies in parameter estimates across measurement errors structures. As expected, given the percentage of non overlapping observations outlined above, the size of the discrepancies is minimum for the first two columns, it increases when one compares columns 1 and 3, and it is maximum when considering the last two columns. It is also not surprising that the parameters most af-

Table 2.2: Estimates of the parameters for the CEV-J option pricing model on the sample of options on the S&P 500 index observed on each day from Jan. 4, 1996 to Dec. 30, 2005. Each column contains the estimates obtained by selecting the NPs to be inverted following one of the five criteria detailed in section 2.3.3. Asymptotic standard errors in parentheses.

	$\tau_* = 30$ $X_* = 0\%$	$\tau_* = 15$ $X_* = 0\%$	$\tau_* = 60$ $X_* = 0\%$	$\tau_* = 30$ $X_* = -1\%$	$\tau_* = 30$ $X_* = 1\%$
λ	-0.98	-1.34	-1.16	-2.01	2.33
	(2.52)	(2.55)	(2.50)	(2.54)	(2.32)
α	2.57×10^{-3}	1.93×10^{-3}	2.27×10^{-3}	5.69×10^{-3}	2.59×10^{-3}
	(7.49×10^{-4})	(6.22×10^{-4})	(1.12×10^{-3})	(2.28×10^{-3})	(6.09×10^{-4})
β	2.72×10^{-2}	2.80×10^{-2}	2.45×10^{-2}	2.12×10^{-2}	2.57×10^{-2}
	(1.03×10^{-2})	(7.11×10^{-3})	(9.94×10^{-3})	(1.76×10^{-2})	(6.38×10^{-3})
γ	0.27	0.27	0.26	0.26	0.29
	(8.41×10^{-4})	(8.09×10^{-4})	(1.03×10^{-3})	(5.69×10^{-4})	(5.75×10^{-4})
φ	1.20	1.21	1.21	1.16	1.22
	(6.49×10^{-4})	(6.17×10^{-4})	(1.03×10^{-3})	(5.61×10^{-4})	(5.41×10^{-4})
ρ	-0.77	-0.79	-0.73	-0.77	-0.93
	(1.79×10^{-3})	(2.16×10^{-3})	(2.76×10^{-3})	(1.44×10^{-3})	(1.41×10^{-3})
$\alpha^{\mathbb{Q}}$	-9.21×10^{-4}	-9.81×10^{-4}	-6.20×10^{-4}	-4.42×10^{-4}	-6.05×10^{-4}
	(1.57×10^{-5})	(2.15×10^{-5})	(1.88×10^{-5})	(2.32×10^{-5})	(1.31×10^{-5})
$\beta^{\mathbb{Q}}$	4.26×10^{-2}	4.26×10^{-2}	4.33×10^{-2}	3.16×10^{-2}	7.02×10^{-2}
	(1.68×10^{-4})	(1.68×10^{-4})	(1.98×10^{-4})	(1.51×10^{-4})	(1.78×10^{-4})
$\lambda_J^{\mathbb{Q}}$	4.22×10^{-2}	4.64×10^{-2}	1.59×10^{-2}	5.40×10^{-2}	0.20
	(3.90×10^{-4})	(4.44×10^{-4})	(7.41×10^{-5})	(4.18×10^{-4})	(1.16×10^{-3})
$\mu_J^{\mathbb{Q}}$	-0.74	-0.69	-1.48	-0.70	8.77×10^{-2}
	(5.77×10^{-3})	(5.99×10^{-3})	(9.59×10^{-3})	(3.05×10^{-3})	(1.68×10^{-3})
λ_J	0.44	0.50	0.29	0.58	2.38
	(2.54×10^{-2})	(2.73×10^{-2})	(1.82×10^{-2})	(4.20×10^{-2})	(0.16)
μ_J	-5.43×10^{-2}	-4.71×10^{-2}	-4.00×10^{-2}	-3.08×10^{-2}	3.95×10^{-3}
	(0.11)	(9.87×10^{-2})	(0.22)	(5.98×10^{-2})	(3.19×10^{-2})
σ_J	1.72	1.69	2.56	0.99	0.91
	(1.52×10^{-2})	(1.46×10^{-2})	(9.00×10^{-3})	(3.17×10^{-3})	(2.07×10^{-3})
loglik.	31579.9	30600.7	26681.1	32691.9	28037.1

fected by the assumption about the measurement error structure are those characterizing the risk-neutral measure, i.e. the diffusion coefficient parameters (with the lone exception of φ), the risk-neutral drift and jump process.

The impact of the measurement error structure on the model's implications can also be evaluated by looking at the filtered volatility trajectories and the option pricing residuals. For simplicity, we focus here on the LOG-J model (analogous results for the CEV-J

Table 2.3: Summary statistics of the filtered volatilities for the LOG-J option pricing model estimated on the sample of options on the S&P 500 index observed on each day from Jan. 4, 1996 to Dec. 30, 2005. $q(p)$ denotes the p-th percentile.

	Volatilities ($\times 10^4$)					Percentage differences w.r.t. $\tau_* = 30$, $X_* = 0\%$			
τ_*	30	15	60	30	30	15	60	30	30
X_*	0%	0%	0%	-1%	1%	0%	0%	-1%	1%
Avg.	0.941	0.925	0.978	0.917	0.546	-10.0	9.9	-0.7	-57.9
Std.Dev.	0.913	0.989	0.915	0.851	0.824	12.7	19.4	13.5	18.0
RMSE	na	na	na	na	na	16.2	21.8	13.6	60.6
Min	0.027	0.013	0.047	0.020	0.003	-50.8	-75.7	-47.8	-90.6
q(0.05)	0.102	0.069	0.129	0.094	0.021	-33.8	-13.1	-15.4	-80.6
q(0.25)	0.318	0.263	0.354	0.320	0.094	-18.0	-3.0	-7.2	-71.7
Median	0.702	0.642	0.737	0.715	0.273	-7.4	4.6	-2.3	-60.6
q(0.75)	1.173	1.147	1.219	1.157	0.606	-1.2	19.5	4.2	-48.0
q(0.95)	2.722	2.866	2.798	2.618	2.065	5.9	48.3	17.8	-21.8
Max	6.705	7.638	6.828	6.588	8.247	143.8	152.7	280.5	23.6

model are available at request). Table 2.3 reports summary statistics computed on the filtered volatilities, along with their percentage differences with respect to the volatility filtered using the first criterion. It is apparent that the specific assumption about the measurement error structure can be quite relevant, and somewhat systematic. For example, the volatilities derived from out-of-the-money (in-the-money) options tend to be systematically higher (lower) than those derived under the baseline criterion; nevertheless, some notable exceptions can be spotted at dates in which the volatility marks (according to some criterion) a sharp increase.

Table 2.4 reports the average option pricing residuals by discounted moneyness X and time to maturity τ. Overall, the model tends to systematically underprice (overprice) short (long) maturity options, irrespectively of the assumption about τ_* and X_*. However, it is apparent that the extent of the pricing errors depend on the characteristics of the contract used to infer the volatility status. For example, the overpricing in medium-to-long maturity, in-the money options almost vanishes when using criterion 5, replaced by a symmetric overpricing for out-of-the-money contracts. Using criterion 3 generates larger underpricing errors for short maturity options. In general, moving from one criterion to

Table 2.4: Average option pricing residuals ($\times 10^4$) by discounted moneyness and maturity, for the LOG-J model and alternative assumptions about τ_* and X_*.

Time to Maturity τ	Discounted Moneyness X					
	(-5%,-3%)	(-3%,-1%)	(-1%,1%)	(1%,3%)	(3%,5%)	All
			$\tau_* = 30,\ X_* = 0\%$			
(15,24)	4.7	0.6	0.2	3.0	4.7	2.4
(25,33)	0.2	-0.9	0.3	0.9	-0.6	0.0
(34,42)	0.3	-0.7	-0.1	-2.5	-6.0	-1.3
(43,51)	-0.5	-1.5	-0.9	-4.4	-8.7	-2.3
(52,60)	-3.3	-3.2	-2.4	-8.0	-12.9	-4.6
All	1.0	-0.8	-0.5	-0.9	-2.1	-0.5
			$\tau_* = 15,\ X_* = 0\%$			
(15,24)	4.4	0.5	-0.2	2.1	3.6	2.0
(25,33)	-0.2	-1.1	0.0	0.4	-1.1	-0.4
(34,42)	0.2	-0.4	0.3	-2.0	-5.5	-1.1
(43,51)	-0.3	-1.3	-0.6	-3.4	-7.4	-1.8
(52,60)	-3.2	-2.9	-1.8	-6.4	-10.6	-3.9
All	0.9	-0.8	-0.5	-0.8	-2.1	-0.5
			$\tau_* = 60,\ X_* = 0\%$			
(15,24)	4.2	1.2	2.8	6.5	8.0	3.9
(25,33)	-0.2	-0.3	2.3	4.0	3.0	1.5
(34,42)	1.0	0.9	2.4	1.2	-1.0	1.1
(43,51)	-0.3	-0.3	0.4	-2.2	-5.0	-1.0
(52,60)	-3.7	-3.6	-4.0	-9.0	-12.5	-5.6
All	0.9	-0.1	1.6	1.9	1.2	1.0
			$\tau_* = 30,\ X_* = -1\%$			
(15,24)	3.1	-0.7	-2.3	-0.1	3.2	0.4
(25,33)	-0.4	-1.1	-0.7	-0.3	0.0	-0.6
(34,42)	0.9	0.7	0.2	-1.6	-2.8	-0.2
(43,51)	-0.2	-0.5	-0.2	-2.8	-4.5	-1.1
(52,60)	-2.6	-1.6	-0.3	-5.2	-6.1	-2.4
All	0.6	-0.7	-0.9	-1.4	-0.7	-0.6
			$\tau_* = 30,\ X_* = +1\%$			
(15,24)	6.7	2.1	0.9	4.4	5.1	3.6
(25,33)	0.5	-1.8	-0.9	1.6	1.3	-0.1
(34,42)	-0.7	-2.7	-1.9	-0.4	-1.2	-1.5
(43,51)	-2.1	-3.6	-1.5	-0.4	-0.2	-1.9
(52,60)	-3.6	-3.2	0.3	1.3	3.3	-1.0
All	1.3	-1.2	-0.4	1.9	2.2	0.4

another tends to reduce the errors for contracts close to τ_* and X_*, and to increase those for contracts with very different characteristics.

Overall, these results suggest that the specific assumption about the measurement error structure is critical both on parameter estimates, volatility filtering and option pricing. We now turn our attention to an alternative inference strategy.

2.4 Estimation by nonlinear filtering

In this section, we illustrate an alternative approach which does not require to assume the existence of an option observed without measurement error. Assuming that each observed NP is affected by observation noise is less arbitrary, and seems more natural, but it complicates the evaluation of the sample loglikelihood, because when there are more sources of uncertainty than observed quantities, the likelihood can not be computed using the Jacobian formula, but requires the evaluation of a high-dimensional integral. In some special cases, e.g. affine term structure models, the problem can be simplified by casting it in a Gaussian state space model and exploiting the Kalman filter recursions, as in De Jong (2000), but in general its solution requires Importance Sampling (IS) techniques. This is the avenue followed e.g. by Brandt and He (2005) in their analysis of affine term structure models.

In this paper, we show how to evaluate the loglikelihood by combining an IS scheme and a Particle Filter algorithm, along the lines suggested by Durham and Gallant (2002, sect. 7); we present it in detail in section 2.4.1. In particular, we highlight that including option prices in the observation sample significantly improves the performance of the simulated maximum likelihood (SML) estimator because of the huge amount of information they convey about the latent state variable, i.e. the volatility.

2.4.1 Likelihood evaluation

Let $\mathcal{F}_t$ be the filtration generated by the variables observed up to time t, i.e. $\{P_s,\ s = 0, 1, ..., t\}$, and $\{\mathbf{H}_s = (H_{is}),\ i = 1, ..., N_s;\ s = 0, 1, ..., t\}$. The likelihood function is given by:

$$
\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\psi}) &= \prod_{t=1}^{T} f(\mathbf{H}_t, P_t | \mathcal{F}_{t-1}) \\
&= \prod_{t=1}^{T} \int f(\mathbf{H}_t | P_t, V_t)\, f(P_t, V_t | P_{t-1}, V_{t-1})\, f(V_{t-1} | \mathcal{F}_{t-1})\, d\, V_t\, d\, V_{t-1}. \qquad (2.8)
\end{aligned}
$$

The second equality derives from the Markov property of the diffusion and the independence of measurement errors. The initial condition P_0 is known, and V_0 will be integrated out (see below). Consider the time t contribution to $L(\boldsymbol{\theta}, \boldsymbol{\psi})$:

$$
f(\mathbf{H}_t, P_t | \mathcal{F}_{t-1}) = \int f(\mathbf{H}_t | P_t, V_t)\, f(P_t, V_t | P_{t-1}, V_{t-1})\, f(V_{t-1} | \mathcal{F}_{t-1})\, d\, V_t\, d\, V_{t-1}. \qquad (2.9)
$$

This two-dimensional integral can be interpreted as an expected value with respect to V_t, V_{t-1} under the distribution implicitly defined by the integrand. Its value can be approximated using an IS scheme by specifying a sampling density for the integration variables. However, the transition pdf $f(P_t, V_t | P_{t-1}, V_{t-1})$ is unknown, and must be approximated using the Modified Brownian Bridge (MBB) strategy outlined in Appendix 2.B. Luckily, the two IS schemes can be merged into a single one. To see how, let $t - 1 = \tau_0 < \tau_1 < ... < \tau_M = t$, and $\boldsymbol{V} = (V_{\tau_1}, \dots, V_{\tau_{M-1}})'$. Following (B-1), the integral on the rhs of (2.9) can be approximated with:

$$
\int f(\mathbf{H}_t | P_t, V_t)\, f^a(P_t | P_{t-1}, V_t, \boldsymbol{V}, V_{t-1})\, f^a(V_t, \boldsymbol{V} | V_{t-1}) f(V_{t-1} | \mathcal{F}_{t-1})\, d\, V_t\, d\, \boldsymbol{V}\, d\, V_{t-1}. \quad (2.10)
$$

where $f^a(P_t | P_{t-1}, V_t, \boldsymbol{V}, V_{t-1})$ and $f^a(V_t, \boldsymbol{V} | V_{t-1})$ are defined in Appendix 2.B. According to (2.10), the likelihood evaluation requires to numerically approximate T integrals whose dimension equals $M + 1$. This can be done using a single IS scheme. Let $q(V_t, \boldsymbol{V} | V_{t-1})$

be a pdf on $\mathbb{R}^M$, and rewrite (2.10) as follows:

$$\int \frac{f(\mathbf{H}_t|P_t, V_t)\, f^a(P_t|P_{t-1}, V_t, \boldsymbol{V}, V_{t-1})\, f^a(V_t, \boldsymbol{V}|V_{t-1})}{q(V_t, \boldsymbol{V}|V_{t-1})}\, q(V_t, \boldsymbol{V}|V_{t-1})\, f(V_{t-1}|\mathcal{F}_{t-1})\, d\,V_t\, d\,\boldsymbol{V}\, d\,V_{t-1}. \tag{2.11}$$

(2.11) highlights that $f(\mathbf{H}_t, P_t|\mathcal{F}_{t-1})$ can be seen as the expected value with respect to $V_t, \boldsymbol{V}, V_{t-1}$ of the ratio in the integrand under the joint distribution defined by the product $q(V_t, \boldsymbol{V}|V_{t-1})\, f(V_{t-1}|\mathcal{F}_{t-1})$. Let $(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l, \widetilde{V}_{t-1}^l)$ be L independent draws from $q(V_t, \boldsymbol{V}|V_{t-1})\, f(V_{t-1}|\mathcal{F}_{t-1})$. The IS estimate of (2.11) is given by:

$$\widetilde{f}^{(L)}(\mathbf{H}_t, P_t|\mathcal{F}_{t-1}) = \frac{1}{L}\sum_{l=1}^{L} \frac{f(\mathbf{H}_t|P_t, \widetilde{V}_t^l)\, f^a(P_t|P_{t-1}, \widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l, \widetilde{V}_{t-1}^l)\, f^a(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l|\widetilde{V}_{t-1}^l)}{q(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l|\widetilde{V}_{t-1}^l)}. \tag{2.12}$$

To implement (2.12), we need to specify (i) which density to choose as $q(V_t, \boldsymbol{V}|V_{t-1})$, and how to draw from it; and (ii) how to draw from $f(V_{t-1}|\mathcal{F}_{t-1})$. The next sections consider these points in turn.

2.4.1.1 The auxiliary density $q(V_t, \boldsymbol{V}|V_{t-1})$

Let V_{t-1} be a generic lagged value of the latent volatility drawn from $f(V_{t-1}|\mathcal{F}_{t-1})$ (we will show how to simulate from this distribution in section 2.4.1.2). In this section, we propose a sampling density $q(V_t, \boldsymbol{V}|V_{t-1})$ for $V_t, \boldsymbol{V}$ which has a simple functional form, is easy to sample from, and provides accurate estimates of the likelihood.

To keep low the MC variance of (2.12), the sampling density $q(V_t, \boldsymbol{V}|V_{t-1})$ should be as much as possible proportional to $f(\mathbf{H}_t|P_t, V_t)\, f^a(P_t|P_{t-1}, V_t, \boldsymbol{V}, V_{t-1})\, f^a(V_t, \boldsymbol{V}|V_{t-1})$ over the whole support of V_t and $\boldsymbol{V}$. This product is informative about the uncertainty surrounding V_t and $\boldsymbol{V}$ in two ways: it reflects (i) the information about V_t in the observed cross-section of NPs $\mathbf{H}_t$ through the measurement error density $f(\mathbf{H}_t|P_t, V_t)$, and (ii) the information about both V_t and $\boldsymbol{V}$ contained in $f^a(P_t, V_t, \boldsymbol{V}|P_{t-1}, V_{t-1})$. In our framework, the second source of information is clearly dominated by the information in the option prices, and this remark suggests that, instead of the usual recursive factorization, it is more convenient to factorize the auxiliary sampling density as:

$$q(V_t, \boldsymbol{V}|V_{t-1}) = q(V_t|V_{t-1})\, q(\boldsymbol{V}|V_t, V_{t-1}).$$

Consider first $q(V_t|V_{t-1})$. Ideally, this density should equal $f^a(V_t|\mathbf{H}_t, P_t, P_{t-1}, V_{t-1})$, which is unavailable, but can be approximated by noting that:

$$f^a(V_t|\mathbf{H}_t, P_t, P_{t-1}, V_{t-1}) \propto f(\mathbf{H}_t|P_t, V_t)\, f^a(P_t, V_t|P_{t-1}, V_{t-1}), \tag{2.13}$$

where the two densities on the rhs correspond to the two sources of information about V_t discussed above.

In this paper, we use as $q(V_t|V_{t-1})$ the Laplace approximation to (2.13). The Laplace approximation is a powerful and accurate strategy widely used in mathematics and statistics to represent unknown densities; see Gelman et al. (1995) for a general presentation, and Durham (2006) and Huber at al. (2009) for two applications in financial econometrics. In a nutshell, it consists of a Gaussian pdf centered at the mode of the target density, with dispersion given by minus the inverse of the Hessian matrix of the log of target, evaluated at the mode. In practice, we proceed as follows. Let us approximate $f^a(P_t, V_t|P_{t-1}, V_{t-1})$ with the Gaussian distribution derived from the Euler discretization over the whole interval $(t-1, t)$ – i.e., ignoring the subintervals defined above. We first compute

$$\hat{V}_t = \arg\max_{V_t} \left[\log f(\mathbf{H}_t|P_t, V_t) + \log f^a(P_t, V_t|P_{t-1}, V_{t-1})\right]$$

using Newton's Method, and

$$\widehat{\Upsilon}_t = \frac{\partial^2}{\partial V_t^2}\left[\log f(\mathbf{H}_t|P_t, \hat{V}_t) + \log f^a(P_t, \hat{V}_t|P_{t-1}, V_{t-1})\right].$$

The Laplace sampling density for V_t is then Gaussian, with mean $\hat{V}_t$ and variance $-1/\widehat{\Upsilon}_t$. Notice that both $\hat{V}_t$ and $\widehat{\Upsilon}_t$ depend on V_{t-1}. In practice, this implies that the Laplace approximation must be computed for each simulated value of the lagged volatility. While this might seem complicated, it should be noted that the whole procedure amounts to solve a large number of straightforward univariate maximization problems, given the availability of good initial points and of analytical expressions of the derivatives of the function to be maximized. Usually (see e.g. Durham, 2006, 2007), the Laplace approximation is computed with respect to the whole trajectory of the volatility state because the likelihood is not sequentially factorized as in (2.8), but rather defined as a single integral with respect to the volatility trajectory, whose dimension is equal to T. In this paper, we prefer to work with the factorized loglikelihood for several reasons. The whole-trajectory strategy is well-suited for discrete-time models, but becomes much more complicated in a continuous-time setting, in which there are multiple "intermediate" volatility values to integrate out. Moreover, the sequential strategy naturally provides a way to compute the generalized residuals that we will use later to conduct a specification analysis.

A couple of remarks about this result are in order. First, the usefulness of our sampling density depends on the validity of two simplifying approximations: using the Euler discretization instead of the true transition density to derive $\hat{V}_t$ and $\widehat{\Upsilon}_t$, assuming at most one jump between $t-1$ and t. These steps, however, can be easily checked ex post by examining the MC variance of (2.12), and checking that this estimate has finite variance. We show in section 2.5 that this variance is actually very low in all our applications.

Second, a similar approach could be used also to approximate the pdf $f(V_0|\mathcal{F}_0) = f(V_0|\mathbf{H}_0, P_0)$ which is needed in order to integrate out V_0 in (2.11) for $t = 1$. To this end, given the lack of a lagged volatility, we use a Gaussian density computed as the Laplace approximation above, but based on a target density that neglects the transition density, and focuses exclusively on the measurement errors density $f(\mathbf{H}_0|P_0, V_0)$.

It remains to discuss our choice of $q(\boldsymbol{V}|V_t, V_{t-1})$, which is a pdf for the "intermediate" volatility states $\boldsymbol{V}$ given V_{t-1} and V_t. The ideal pdf would be $f^a(\boldsymbol{V}|\mathbf{H}_t, P_t, V_t, P_{t-1}, V_{t-1})$, which is unknown. However, given the density used to draw V_t, we argue that no information about $\boldsymbol{V}$ is lost if we drop the conditioning on $\mathbf{H}_t$, P_t and P_{t-1}. This allows to factorize $q(\boldsymbol{V}|V_t, V_{t-1})$ as:

$$q(\boldsymbol{V}|V_t, V_{t-1}) = \prod_{m=1}^{M-1} q(V_m|V_t, V_{m-1}).$$

We set each pdf in the product of the rhs as a Gaussian density with moments computed in the same way as the MBB strategy discussed in Appendix 2.B. Notice however that, unlike the "pure " MBB approach, the simulated $\boldsymbol{V}$ trajectories do not start from the same volatility state V_{t-1}, and do not end up in the same volatility state V_t, as both these values are simulated by $f(V_{t-1}|\mathcal{F}_{t-1})$ and $q(V_t|V_{t-1})$, respectively.

2.4.1.2 Drawing from $f(V_t|\mathcal{F}_t)$

In this section, we show how to randomly draw from $f(V_t|\mathcal{F}_t)$ for $t \geq 1$; the case $t = 0$ was already discussed in section 2.4.1.1. The approach we adopt is basically the application of the Particle Filter discussed by Durham and Gallant (2002, sect. 7).

Let $\{(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l, \widetilde{V}_{t-1}^l),\ l = 1, ..., L\}$ be L independent draws from $q(V_t, \boldsymbol{V}|V_{t-1})\, f(V_{t-1}|\mathcal{F}_{t-1})$, and

$$\eta_t^l = \frac{f(\mathbf{H}_t|P_t, \widetilde{V}_t^l)\, f^a(P_t|P_{t-1}, \widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l, \widetilde{V}_{t-1}^l)\, f^a(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l|\widetilde{V}_{t-1}^l)}{q(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l|\widetilde{V}_{t-1}^l)}, \quad l = 1, ..., L$$

be the L simulated values of the ratio under integration used in (2.12). Define the normalized weights:

$$\overline{\eta}_t^l = \frac{\eta_t^l}{\sum_{k=1}^L \eta_t^k}.$$

By construction, $\overline{\eta}_t^l \in (0, 1)$ and $\sum_{l=1}^{L} \overline{\eta}_t^l = 1$. By Theorem 1 in Geweke (1989), the collection $\{[(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l, \widetilde{V}_{t-1}^l), \overline{\eta}_t^l],\ l = 1, ..., L\}$ can be seen as a discrete approximation of $f(V_t, \boldsymbol{V}, V_{t-1}|\mathcal{F}_t)$, in the sense that, by a Law of Large Numbers:

$$\sum_{l=1}^{L} \overline{\eta}_t^l\, g(\widetilde{V}_t^l, \widetilde{\boldsymbol{V}}^l, \widetilde{V}_{t-1}^l) \quad \xrightarrow{p} \quad \int g(V_t, \boldsymbol{V}, V_{t-1})\, f(V_t, \boldsymbol{V}, V_{t-1}|\mathcal{F}_t)\, dV_t\, d\boldsymbol{V}\, dV_{t-1},$$

for any function g for which the expectation on the rhs exists and is finite.

There are various different ways to exploit this result to draw V_t from $f(V_t|\mathcal{F}_t)$. The simplest one, advanced by Rubin (1988), consists in drawing with replacement from $\{\widetilde{V}_t^l,\ l = 1, ..., L\}$, where $\overline{\eta}_t^l$ is the probability that each $\widetilde{V}_t^l$ is drawn. The resulting likelihood, however, would not be continuous in the parameters, causing difficulties to the numerical optimizer. Durham and Gallant (2002, sect. 7) prefer to use the collection $\{(\widetilde{V}_t^l, \overline{\eta}_t^l),\ l = 1, ..., L\}$ to build a Hermite approximation of $f(V_t|\mathcal{F}_t)$, and draw from it. In this paper, we use the bootstrap procedure based on univariate linear spline advanced by Pitt (2002) to get an approximated likelihood which is smooth in the parameters. With a multivariate latent state some other kind of multivariate interpolation technique should be used.

2.4.2 Diagnostic testing and filtered (generalized) residuals

Given the ML estimates of the parameters, we use simulation based techniques to estimate sequences of filtered estimates of the latent volatility V_t and of functions of V_t. These estimates can be used to assess the validity of the models specification and to get some intuition about their deficiencies.

We consider filtering with respect to 4 alternative kinds of information sets. The first one is $\mathcal{F}_{t-1}$, and it contains, at date t, only the observations up to date $t-1$. Conditioning on this information set, the filtered estimate of a generic function $\zeta(\cdot, \cdot)$ is defined by:

$$
\begin{aligned}
\mathrm{E}[\zeta(P_t, V_t)|\mathcal{F}_{t-1}] &= \int \zeta(P_t, V_t) f(P_t, V_t|\mathcal{F}_{t-1}) dP_t\, dV_t \\
&= \int \zeta(P_t, V_t) f(P_t|P_{t-1}, V_t, \boldsymbol{V}, V_{t-1}) f(V_t, \boldsymbol{V}|V_{t-1}) f(V_{t-1}|\mathcal{F}_{t-1}) dP_t\, dV_t\, d\boldsymbol{V}\, dV_{t-1}.
\end{aligned}
$$

To estimate these quantities, we first use the particle filter technique outlined in section 2.4.1.2 to generate draws from $f(V_{t-1}|\mathcal{F}_{t-1})$, and then simulate the couple $(\boldsymbol{V}, V_t)$ from $f(V_t, \boldsymbol{V}|V_{t-1})$ by drawing blindly from the Euler discretization of the V_t process. Finally, P_t is drawn from the conditional distribution $f(P_t|P_{t-1}, V_t, \boldsymbol{V}, V_{t-1})$ derived in Appendix 2.B. We label "predicted" values of $\zeta(P_t, V_t)$ the results of this procedure. They can be useful in diagnostic checking, but they do not represent the best way to predict the price of an option, in particular when some information about the forward date is available.

To this end, we consider some alternative information sets $\mathcal{F}_t^a$ that contain $\mathcal{F}_{t-1}$, P_t and a subset $\mathbf{H}_t^a$ of the whole set of options $\mathbf{H}_t$ observed at date t. In this case, the filtered value of a generic function of the volatility V_t (notice that since P_t is known conditionally on $\mathcal{F}_t^a$, we can omit it from the arguments of ζ) is defined by:

$$\begin{aligned}
\mathrm{E}[\zeta(V_t)|\mathcal{F}_t^a] &= \int \zeta(V_t) f(V_t|\mathcal{F}_t^a) dV_t \\
&= \frac{\int \zeta(V_t) f(\mathbf{H}_t^a|P_t, V_t) f(P_t|P_{t-1}, V_t, \boldsymbol{V}, V_{t-1}) f(V_t, \boldsymbol{V}|V_{t-1}) f(V_{t-1}|\mathcal{F}_{t-1})\, dV_t\, d\boldsymbol{V}\, dV_{t-1}}{\int f(\mathbf{H}_t^a|P_t, V_t) f(P_t|P_{t-1}, V_t, \boldsymbol{V}, V_{t-1}) f(V_t, \boldsymbol{V}|V_{t-1}) f(V_{t-1}|\mathcal{F}_{t-1})\, dV_t\, d\boldsymbol{V}\, dV_{t-1}}
\end{aligned}$$

These integrals can be evaluated following the procedure described in section 2.4.1. Notice that the denominator is equivalent to (2.10) using only the subset $\mathbf{H}_t^a$ of options instead of the full vector $\mathbf{H}_t$; if $\mathcal{F}_t^a$ does not contain options at date t, then the $f(\mathbf{H}_t^a|P_t, V_t)$ factor disappears from both integrals. To evaluate the numerator, we use the same simulated values of the denominator integrand, multiplied by $\zeta(\cdot)$ evaluated at the simulated V_t value. Geweke (1989) shows that the simulation variance of the estimate of the ratio is reduced if the same draws are used in the numerator and the denominator.

We consider three kinds of augmented information sets: (i) $\mathcal{F}_t^P = (P_t, \mathcal{F}_{t-1})$; (ii) $\mathcal{F}_t^* = (P_t, H_t^*, \mathcal{F}_{t-1})$, where H_t^* is the date t option selected according to criterion 1, as defined in section 2.3.3; and finally, (iii) $\mathcal{F}_t = (P_t, \mathbf{H}_t, \mathcal{F}_{t-1})$. Case (i) considers the predicted values of $\zeta(V_t)$ given past observations and the current value of the stock index. Knowledge of the latter carries some information on the current V_t value, and should allow more accurate predictions. We label "updated given P_t" the results of this procedure.

Case (ii) further enlarges the information set by including a single option at each date. Since the latter is a monotone function of V_t, this inclusion significantly increases the available information on the latent variable, and generates dramatically improved predictions of $\zeta(V_t)$, that we label "updated given P_t and H_t^*". This kind of conditioning is very interesting from an operational point of view, as it allows to price any date t option conditionally on the observed price H_t^*.

It should be noted that pricing some options relative to one observed option is also possible in the approach of section 2.3, in which one option at each date was assumed to be free of measurement error. The two procedures, however, are fundamentally different. On one hand, volatility is filtered by inverting an observed option price; on the other, it is filtered by estimating a conditional expectation given an option which is observed with error. It is likely that, if such error actually exists, neglecting it might induce biased volatility and option prices estimates. Furthermore, in our approach, there is no need to condition on just one option at each date; we might as well condition on all but one of the observed options, and compute the predicted price of the contract left over. This should further enhance the accuracy of the predictions, and it is of course impossible to do under the assumptions of the approach of section 2.3.

Case (*iii*) considers the widest information set comprising the log stock price and all the options observed at each date. We label the values of $\zeta(V_t)$ predicted in this way as "fully updated". Notice that their computation can be done using the volatility trajectories used in the likelihood evaluation discussed in section 2.4.1. In all cases, 100,000 trajectories were used to approximate the above expressions using Monte Carlo integration techniques.

In our set up, predicted values can be computed for the options NPs and the log stock index price. In the case of options, they allow to compute residuals that, according to our hypothesis about measurement errors, should conform to a Gaussian distribution independent across dates. This can be checked using standard test procedures, such as the Box-Pierce test, either applied to the residuals or to their squares, and the Jarque-Bera test.

In the case of the log stock prices, the assumed distribution is not Gaussian; rather, it is a mixture of conditionally heteroskedastic Gaussian densities. To perform diagnostic checking, we computed the associated generalized residuals as follows. Consider the first kind of filtering rule discussed above, based on the conditioning information set $\mathcal{F}_{t-1}$, and the predicted values for $\zeta(P_t, V_t) = F(P_t|P_{t-1}, V_t, V_{t-1})$, the conditional cdf corresponding to the pdf derived in Appendix 2.B. If the model specification is correct, these predicted values should be IID uniformly distributed in [0,1]. If we further transform these uniform generalized residuals using the inverse standard Gaussian cdf, we obtain generalized residuals that, under the null hypothesis of correct specification, should be IID standard Gaussian. This can be tested using, as in the case of options, the Box-Pierce or the Jarque-Bera tests.

2.5 Implementing the SML estimator

2.5.1 Parameter estimation

We estimated a few variants of the LOG-J and the CEV-J models, differing by the parameterization of the volatility and the jump risk premia, along with the corresponding affine specifications. We limit our discussion to the two versions of the LOG-J and the CEV-J models, which, according to information criteria, should be preferred. The "baseline" LOG-J and CEV-J models coincide with the specification estimated in section 2.3.3; the "augmented" versions labelled LOG-Ja and CEV-Ja assume different jump size variances under the actual and the risk-neutral measures.

Table 2.5 reports the results. For each parameter, the table contains the ML estimate, the estimated asymptotic standard error (in parentheses), and the numerical standard error (in brackets). The latter is also computed for the loglikelihood at the optimum, and it is based on 100 estimates using independent draws. The R^2 coefficients of the polynomial interpolation that we use to approximate option prices are equal to 0.9998 for each model, and suggest that the approximation is reliable.

The comparison between the asymptotic and the numerical standard errors weights the relative importance of the "statistical" sampling uncertainty vs. the "numerical" one induced by simulations. The latter is smaller than the former by at least two orders of

Table 2.5: Estimates of parameters for the LOG-J and CEV-J option pricing models on the sample of options on the S&P 500 index observed on each day from Jan. 4, 1996 to Dec. 30, 2005. Asymptotic (statistical) standard errors in parentheses; numerical (Monte Carlo) standard errors in brackets.

	CEV-J	CEV-Ja	LOG-J	LOG-Ja
λ	-3.68	-2.29	1.19	4.86
	(2.15)	(2.86)	(2.47)	(2.59)
	[1.29×10^{-2}]	[1.21×10^{-2}]	[8.15×10^{-3}]	[3.40×10^{-2}]
α	1.43×10^{-3}	8.15×10^{-4}	-2.45×10^{-2}	-2.07×10^{-2}
	(5.45×10^{-4})	(4.59×10^{-4})	(8.33×10^{-3})	(7.30×10^{-3})
	[8.99×10^{-6}]	[1.11×10^{-5}]	[2.19×10^{-5}]	[3.72×10^{-5}]
β	3.08×10^{-2}	3.27×10^{-2}	-3.23×10^{-2}	-2.74×10^{-2}
	(2.69×10^{-2})	(3.16×10^{-2})	(9.32×10^{-3})	(8.80×10^{-3})
	[3.98×10^{-5}]	[2.95×10^{-5}]	[4.88×10^{-5}]	[4.97×10^{-5}]
γ	0.27	0.27	0.29	0.27
	(8.64×10^{-4})	(9.55×10^{-4})	(2.24×10^{-3})	(1.73×10^{-3})
	[6.01×10^{-6}]	[3.43×10^{-6}]	[7.56×10^{-6}]	[9.00×10^{-6}]
ϕ	1.08	1.09	n.a.	n.a.
	(3.62×10^{-4})	(4.26×10^{-4})		
	[2.73×10^{-5}]	[2.81×10^{-6}]		
ρ	-0.75	-0.77	-0.80	-0.85
	(2.06×10^{-3})	(2.28×10^{-3})	(4.29×10^{-3})	(3.91×10^{-3})
	[2.44×10^{-5}]	[1.76×10^{-5}]	[1.81×10^{-5}]	[2.69×10^{-5}]
λ_J	0.17	1.34	0.28	1.68
	(1.27×10^{-2})	(0.16)	(1.68×10^{-2})	(0.14)
	[1.36×10^{-4}]	[2.96×10^{-4}]	[8.55×10^{-5}]	[4.61×10^{-4}]
μ_J	-0.23	-1.10×10^{-2}	-1.75×10^{-2}	-2.20×10^{-2}
	(5.89×10^{-2})	(1.83×10^{-2})	(0.25)	(9.63×10^{-3})
	[1.06×10^{-3}]	[7.77×10^{-5}]	[1.49×10^{-4}]	[2.45×10^{-4}]
σ_J	3.59	0.53	3.15	0.53
	(3.13×10^{-2})	(3.28×10^{-2})	(3.27×10^{-2})	(2.01×10^{-2})
	[1.53×10^{-4}]	[9.82×10^{-5}]	[9.59×10^{-5}]	[1.02×10^{-4}]

magnitude; apparently, our IS strategy succeeds in reducing simulation variance to an acceptable level. We also conducted the tests of the null that the IS estimate (2.12) has finite variance proposed by Monahan (1993) and Koopman et al. (2009), which consider an inequality restriction on a parameter of a Generalized Pareto density, and can be based on a nonparametric procedure (Monahan, 1993) or on the trilogy of ML tests (Koopman et al., 2009). In both cases, these tests must be conducted separately for each integral, which means that in our sample the procedure must be repeated $T = 2516$. A joint

Table 2.5: Continued from the previous page.

	CEV-J	CEV-Ja	LOG-J	LOG-Ja
$\alpha^{\mathbb{Q}}$	9.52×10^{-4}	7.89×10^{-4}	-1.40×10^{-2}	-1.09×10^{-2}
	(3.61×10^{-5})	(3.59×10^{-5})	(3.57×10^{-4})	(2.54×10^{-4})
	$[3.79\times10^{-7}]$	$[2.00\times10^{-7}]$	$[1.20\times10^{-6}]$	$[1.28\times10^{-6}]$
$\beta^{\mathbb{Q}}$	2.01×10^{-2}	2.02×10^{-2}	-8.91×10^{-3}	-9.01×10^{-3}
	(1.51×10^{-4})	(1.60×10^{-4})	(4.77×10^{-5})	(4.66×10^{-5})
	$[9.65\times10^{-7}]$	$[5.65\times10^{-7}]$	$[5.29\times10^{-7}]$	$[4.39\times10^{-7}]$
$\lambda_J^{\mathbb{Q}}$	2.05×10^{-2}	1.99×10^{-2}	2.63×10^{-2}	2.22×10^{-2}
	(2.95×10^{-4})	(2.90×10^{-4})	(4.38×10^{-4})	(3.59×10^{-4})
	$[2.05\times10^{-6}]$	$[8.37\times10^{-7}]$	$[1.29\times10^{-6}]$	$[8.14\times10^{-7}]$
$\mu_J^{\mathbb{Q}}$	-1.05	-1.07	-0.79	-0.83
	(8.64×10^{-3})	(8.39×10^{-3})	(7.39×10^{-3})	(7.84×10^{-3})
	$[9.63\times10^{-5}]$	$[5.09\times10^{-5}]$	$[5.37\times10^{-5}]$	$[4.44\times10^{-5}]$
$\sigma_J^{\mathbb{Q}}$	n.a.	3.75	n.a.	3.53
		(3.26×10^{-2})		(3.59×10^{-2})
		$[1.22\times10^{-4}]$		$[9.43\times10^{-5}]$
ψ_0	-2.16	-2.20	-2.12	-2.24
	(2.38×10^{-2})	(2.38×10^{-2})	(2.38×10^{-2})	(2.36×10^{-2})
	$[1.21\times10^{-4}]$	$[1.30\times10^{-4}]$	$[9.13\times10^{-5}]$	$[9.74\times10^{-5}]$
ψ_X	18.90	19.41	18.45	19.06
	(0.26)	(0.27)	(0.26)	(0.26)
	$[2.34\times10^{-3}]$	$[1.38\times10^{-3}]$	$[1.45\times10^{-3}]$	$[1.31\times10^{-3}]$
ψ_{X^2}	5.92×10^{2}	5.93×10^{2}	6.09×10^{2}	6.05×10^{2}
	(8.64)	(8.72)	(8.41)	(8.48)
	$[8.33\times10^{-2}]$	$[5.50\times10^{-2}]$	$[3.99\times10^{-2}]$	$[4.84\times10^{-2}]$
ψ_τ	-85.47	-85.08	-88.98	-86.96
	(0.55)	(0.55)	(0.54)	(0.54)
	$[1.58\times10^{-3}]$	$[1.30\times10^{-3}]$	$[1.05\times10^{-3}]$	$[1.33\times10^{-3}]$
ψ_{τ^2}	4.51×10^{2}	4.51×10^{2}	4.75×10^{2}	4.68×10^{2}
	(2.76)	(2.77)	(2.77)	(2.77)
	$[9.36\times10^{-3}]$	$[6.04\times10^{-3}]$	$[8.71\times10^{-3}]$	$[7.38\times10^{-3}]$
loglik.	40385.3	40840.0	40971.7	41589.0
	[0.30]	[0.31]	[0.12]	[0.17]

statistic can then be formed by aggregating the univariate test results as in Rao (1952, p. 44). To save space, we do not report the results, but they support the null hypothesis of finite variance of the sample weights.

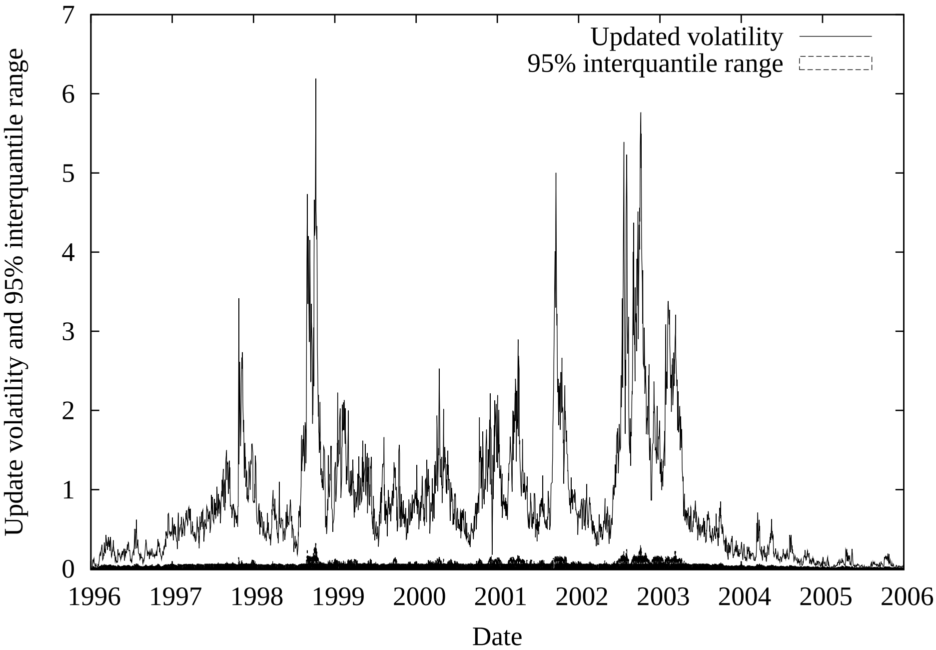
The estimates of the parameters are in line with results reported elsewhere, obtained using different estimation procedures and data. Some differences can be spotted for the parameters that appear both in $\mathbb{P}$ and in $\mathbb{Q}$, such as those in the volatility diffusion coefficient, in the correlation coefficient ρ or in the jump size variance; in general, our estimates are closer to those obtained using only option prices (see, e.g., Bakshi et al., 1997), than to those obtained when the analysis is limited to just one option per day. The reason for this is simply the larger size of the option sample with respect to the stock price sample. The same feature implies that risk neutral parameters are estimated more accurately than their historical counterparts.

With the only exception of λ, all estimates are very accurate and highly significant, in particular for the risk-neutral parameters. This is due to the fact that we completely exploit the large cross-sectional dimension of the sample, instead of focussing only on a single option per day. The estimates imply a stationary volatility process under both $\mathbb{P}$ and $\mathbb{Q}$ and for every specification, which stands in contrast with the nonstationarity of the CEV-J V_t process under $\mathbb{Q}$ suggested by the parameter estimates reported in table 2.2 using the approach of section 2.3.

The estimated jump intensities and jump size distributions are quite different under $\mathbb{P}$ and $\mathbb{Q}$. The comparison between the baseline LOG-J and CEV-J models, and their augmented LOG-Ja and CEV-Ja versions, highlights the importance of allowing different jump size variances under the two measures. The loglikelihood of the augmented specification is higher by 455 points in the CEV-Ja case, and by 618 points in the LOG-Ja case; the constraint $\sigma_J = \sigma_J^{\mathbb{Q}}$ is clearly rejected by a LR test. The augmented specifications suggest that the jumps are very frequent, essentially zero on average with low dispersion under $\mathbb{P}$, and much rarer, but around -1% on average and with a standard deviation of about 3-3.5% under $\mathbb{Q}$. Overall, these results suggest that the the jump component is less important than what is usually acknowledged in previous works, with the partial exception of Durham (2010). This discrepancy could be due to a different sample or model, as non-affine models could capture features of the distribution that a less flexible affine specification would attribute to jumps.

The comparison between the LOG-Ja and the CEV-Ja models is not as simple as that between the baseline and the augmented specifications, since the former are not nested.

Figure 2.1: Average and 95% interquantile range for filtered volatility in the LOG-Ja model for the sample of options on the S&P 500 index, Jan. 4, 1996 to Dec. 30, 2005.



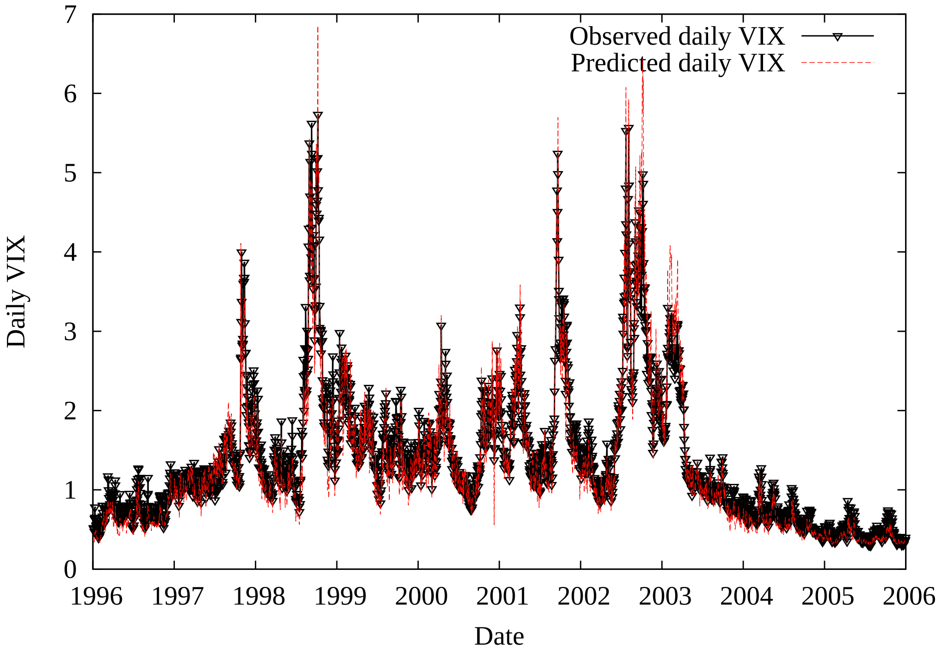
We prefer the log volatility model, not only because of the higher loglikelihood value, but also on the basis of its goodness of fit and of the diagnostic checks based on filtered option prices residuals and log stock price generalized residuals described in section 2.4.2. For this reason, we limit our discussion of analysis of the filtered (generalized) residuals to the LOG-Ja model.

2.5.2 Predicted values and diagnostic tests

Figure 2.1 illustrates the updated latent volatility V_t in the LOG-Ja model. The plot also depicts the 95% interquantile range of the updated volatilities, which witnesses the accuracy of the filtering procedure; actually, the narrowness of the range makes it necessary to plot it as a separate line, instead of a band surrounding the sample average. The shape of the filtered volatility trajectory is in line with similar plots relative to the same period reported elsewhere.

As a way to cross-validate the filtered volatility trajectory, figure 2.2 compares the

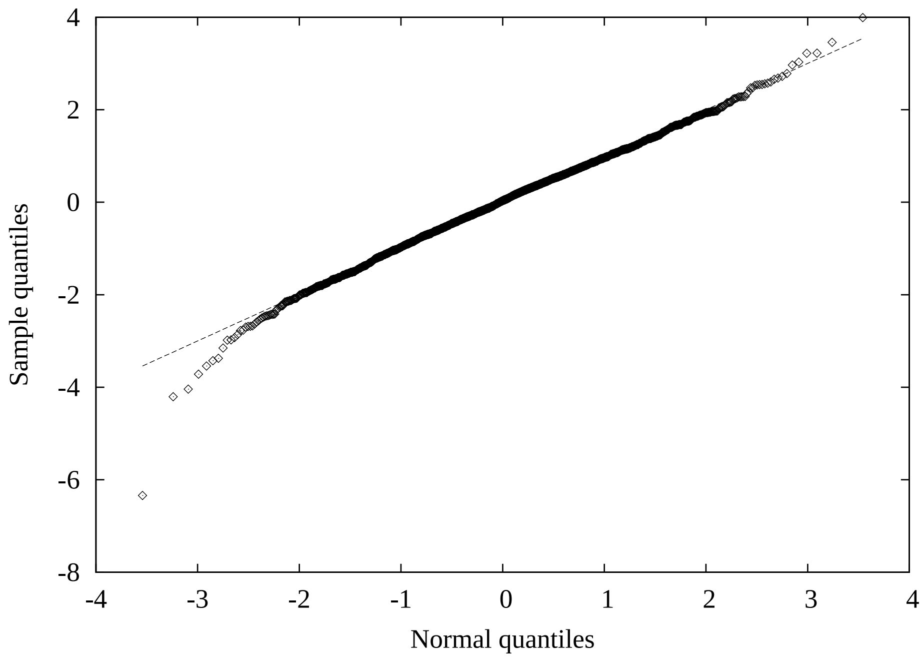
Figure 2.2: Observed daily VIX and predicted daily VIX computed on the basis of the average filtered volatility in the LOG-Ja model for the sample of options on the S&P 500 index, Jan. 4, 1996 to Dec. 30, 2005.



observed VIX contract price over the same period (in daily percentage volatility) with the same quantity implied by the LOG-Ja model and computed numerically as in Jiang and Tian (2005) and Durham (2010). The two trajectories are strikingly similar in shape and very close to each other, although some differences can be spotted at some high-volatility periods: the Asian currency crisis in July 1997 and the mini-crash of October 27, 1997; the LTCM bailout around mid-1998 and the Russian Default in August of the same year; and the period from October 2002 to April 2003, marked by the escalation of the Iraq crisis and the break out of the second Gulf War. In spite of these discrepancies, the correlation between the two series is 0.973.

Inspection of the time series of the predicted generalized residuals of the log index price highlights a few notable outliers, the most striking of which is the large negative value corresponding to the October 27, 1997 mini-crash. This anomaly is shared by all the specifications we examined, and points to the need of a more flexible specification.

Figure 2.3: Gaussian Q-Q plot of the S & P 500 log stock index price predicted generalized residuals from Jan. 4, 1996 to Dec. 30, 2005.



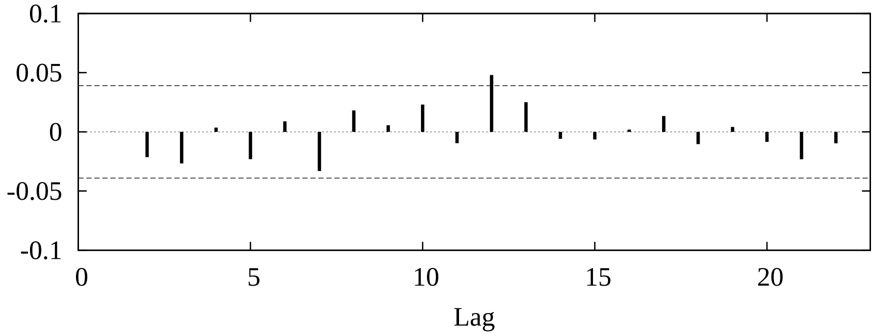
Given the outliers, it is not surprising that a Jarque-Bera test of normality rejects the null overwhelmingly (the test statistic is equal to 73.1, with p-value equal to 6.5×10^{-26}). Figure 2.3 illustrates the Gaussian Q-Q plot of the predicted residuals, in order to get a visual interpretation of this conclusion. The plot clearly points to the left tail as the major source of misspecification.

The correlograms of the predicted residuals and of their squares are illustrated in figure 2.4. The residuals look fairly uncorrelated over time, but not their squares, which is coherent with unexplained conditional heteroskedasticity. A more sophisticated specification of the volatility process – e.g., a two factor volatility model – could help solving this issue.

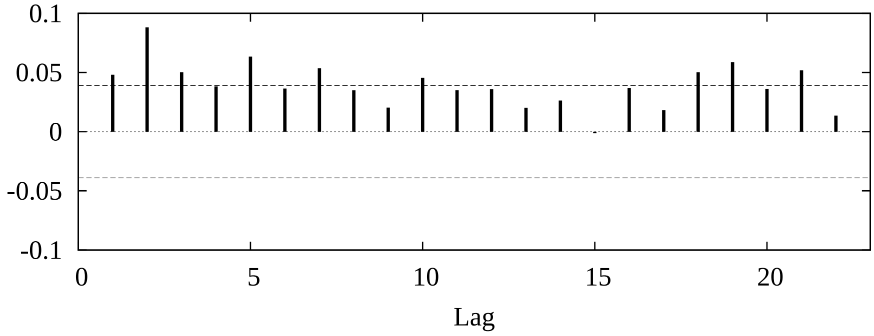
Finally, we examine the various generalized residuals given by the difference between observed option NPs and model implied NPs estimated conditionally on four information sets. The most useful for diagnostic checking are the predicted residuals, computed conditionally on the previous period information set; for pricing purposes, conditioning

Figure 2.4: Autocorrelation functions (with 95% confidence intervals) of the S & P 500 log stock index price predicted generalized residuals from Jan. 4, 1996 to Dec. 30, 2005.

(a) ACF of the predicted generalized residuals



(b) ACF of the squared predicted generalized residuals



on enlarged information sets may be more interesting. Given the wealth of our sample, the descriptive statistics of the options NPs residuals take several tables. Tables 2.6 and 2.7 report the sample averages and standard deviations computed for the residuals conditioned on the four information sets, and disaggregated by time to maturity and discounted moneyness. In general, average residuals are small and close to results reported in other work – see e.g. table 3 of Bates (2000), which is closest in spirit with the fourth panel. Average residuals decrease with the amount of conditioning information (moving from the top to the bottom panel); moreover, the predicted residuals suggest that the model systematically overprices (underprices) out-of-the-money (in-the-money) options, but this bias seems to vanish in the remaining panels.

The effect of an augmented information set is clearer for standard deviations. On average, standard deviations of option residuals decrease by 73% by conditioning on

Table 2.6: Sample averages ($\times 10^4$) of option pricing residuals computed according to different filtering rules by discounted moneyness and maturity for the LOG-Ja model.

Time to Maturity τ	Discounted Moneyness X (-5%,-3%)	(-3%,-1%)	(-1%,1%)	(1%,3%)	(3%,5%)	All
Predicted residuals						
(15,24)	-2.5	-7.2	-5.9	6.4	15.4	-0.9
(25,33)	-5.5	-9.2	-7.9	3.8	8.7	-3.6
(34,42)	-3.8	-8.2	-4.9	3.0	7.9	-2.7
(43,51)	-2.5	-4.6	0.2	7.7	4.5	0.1
(52,60)	-7.2	-8.2	-7.6	1.6	4.6	-5.2
All	-4.2	-7.7	-5.6	4.6	9.4	-2.4
Updated residuals given P_t						
(15,24)	3.0	0.5	0.4	2.9	4.7	1.9
(25,33)	-1.6	-2.2	-1.2	0.3	-0.2	-1.1
(34,42)	-0.1	-0.5	0.3	-0.5	-1.8	-0.3
(43,51)	0.0	0.4	1.8	-0.2	-3.7	0.2
(52,60)	-3.3	-2.6	-1.2	-4.2	-6.7	-3.0
All	-0.1	-0.8	-0.1	0.2	-0.3	-0.3
Updated residuals given P_t and H_t^*						
(15,24)	2.5	0.0	-0.2	2.2	4.0	1.4
(25,33)	-1.8	-2.3	-1.2	-0.5	-1.2	-1.5
(34,42)	-0.9	-1.2	-0.6	-2.1	-3.7	-1.5
(43,51)	-0.9	-0.5	0.8	-2.0	-5.3	-1.0
(52,60)	-4.0	-3.0	-1.8	-5.8	-8.4	-3.8
All	-0.7	-1.3	-0.7	-1.0	-1.6	-1.0
Fully updated residuals						
(15,24)	2.9	0.3	-0.1	1.9	3.5	1.4
(25,33)	-0.2	-0.6	0.5	0.9	0.1	0.1
(34,42)	0.1	-0.1	0.6	-1.1	-2.9	-0.3
(43,51)	-0.2	0.0	0.7	-2.2	-5.4	-0.7
(52,60)	-2.1	-1.1	0.0	-4.6	-7.3	-2.1
All	0.4	-0.3	0.3	-0.4	-1.1	-0.1

the current stock index price in addition to last period's information. This decrease is further enhanced to 81% by adding an option, and reaches 86% for fully updated values. All panels show that pricing errors for options with longer time to maturity are more dispersed; the same is true with respect to discounted moneyness, but this is probably due to higher prices of deep in-the-money options.

Table 2.8 reports sample correlations between contemporaneous option residuals and their squares by time to maturity and discounted moneyness. To save space, we limit the

Table 2.7: Sample standard deviations ($\times 10^4$) of option pricing residuals computed according to different filtering rules by discounted moneyness and maturity for the LOG-Ja model.

Time to Maturity τ	Discounted Moneyness X (-5%,-3%)	(-3%,-1%)	(-1%,1%)	(1%,3%)	(3%,5%)	All
Predicted residuals						
(15,24)	31.2	43.7	60.9	79.7	89.0	60.6
(25,33)	34.5	45.4	62.1	76.3	84.8	60.4
(34,42)	39.1	48.8	62.6	78.7	84.1	61.8
(43,51)	41.6	50.7	61.3	76.3	83.6	60.9
(52,60)	46.9	51.0	63.4	78.6	81.5	62.1
All	37.8	47.3	62.1	78.0	85.4	61.1
Updated residuals given P_t						
(15,24)	10.4	11.9	14.5	16.8	19.0	14.3
(25,33)	11.0	12.7	15.0	17.1	19.6	14.8
(34,42)	11.5	13.1	14.9	17.6	18.6	14.8
(43,51)	13.4	16.1	17.7	20.0	21.4	17.4
(52,60)	16.4	18.3	21.2	24.3	23.6	20.4
All	12.5	14.2	16.5	18.8	20.4	16.1
Updated residuals given P_t and H_t^*						
(15,24)	8.1	7.6	9.2	12.0	15.0	10.2
(25,33)	8.2	6.6	6.8	9.3	15.9	9.2
(34,42)	7.2	5.9	5.8	8.4	11.5	7.6
(43,51)	10.5	11.4	13.1	14.4	15.5	12.8
(52,60)	18.2	15.5	18.0	18.0	20.2	17.8
All	10.9	9.6	11.7	12.5	16.1	11.8
Fully updated residuals						
(15,24)	7.1	6.2	6.8	9.5	13.0	8.3
(25,33)	6.0	4.5	4.6	7.9	12.2	6.9
(34,42)	5.0	3.1	3.3	6.5	10.5	5.6
(43,51)	5.9	4.9	5.5	8.4	11.4	7.0
(52,60)	11.7	10.9	12.3	14.0	16.0	12.7
All	7.6	6.3	7.0	9.6	13.2	8.3

table to the predicted residuals and to three classes of time to maturity and discounted moneyness. Consider the top left entry in the top panel: 0.94 is the sample correlation between pricing residuals observed at each date for short-lived deep out-of-the-money options. Since this computation is not limited to one option per date, the result is not equal to 1, as it is for standard correlation matrices; notice, however, that these matrices are symmetric. Any other entry should be interpreted as the sample correlation between

Table 2.8: Sample contemporaneous correlations of predicted option pricing residuals and squared residuals by discounted moneyness and maturity for the LOG-Ja model. τ_k, $k = h, m, l$ means that time to maturity belongs to (15,30), (31,45), (46,60), resp. X_k, $k = h, m, l$ means that discounted moneyness belongs to (-5%,-1.66%), (-1.66%,1.66%), (1.66%,5%), respectively.

		Time to maturity τ								
		τ_l			τ_m			τ_h		
		Discounted Moneyness X								
τ	X	X_l	X_m	X_h	X_l	X_m	X_h	X_l	X_m	X_h
Residuals										
	X_l	0.94	0.90	0.86	0.88	0.86	0.83	0.91	0.87	0.82
τ_l	X_m		0.97	0.96	0.92	0.96	0.95	0.92	0.95	0.93
	X_h			0.99	0.92	0.97	0.98	0.87	0.95	0.97
	X_l				0.96	0.94	0.90	0.91	0.90	0.90
τ_m	X_m					0.98	0.97	0.93	0.96	0.97
	X_h						1.00	0.91	0.96	0.98
	X_l							0.98	0.95	0.88
τ_h	X_m								0.99	0.97
	X_h									1.00
Squared residuals										
	X_l	0.91	0.83	0.61	0.76	0.74	0.65	0.86	0.76	0.50
τ_l	X_m		0.93	0.78	0.76	0.87	0.83	0.79	0.86	0.63
	X_h			0.96	0.80	0.89	0.94	0.55	0.75	0.91
	X_l				0.93	0.88	0.77	0.85	0.79	0.81
τ_m	X_m					0.96	0.89	0.87	0.90	0.92
	X_h						0.97	0.82	0.90	0.94
	X_l							0.94	0.89	0.54
τ_h	X_m								0.95	0.78
	X_h									0.98

contemporaneous pricing residuals associated to contracts with characteristics belonging to different classes.

The predicted residuals and their squares are very correlated; the lowest entry in the top panel is 0.82, corresponding to contracts with completely opposite characteristics. The correlations are lower for squared predicted residuals, but still very high. A plausible explanation of these results is that the model lacks jumps in volatility. In this case, predictive filtering can not anticipate the occurrence of a jump in the latent state, inducing large price residuals of the same sign, which in turn are reflected in large and positive

Table 2.9: Sample statistics of predicted option pricing residuals by discounted moneyness and maturity for the LOG-Ja model.

Time to Maturity τ	Discounted Moneyness X (-5%,-3%)	(-3%,-1%)	(-1%,1%)	(1%,3%)	(3%,5%)	All
			Skewness			
(15,24)	-2.1	-1.9	-0.6	-0.3	0.0	-1.1
(25,33)	-0.3	-0.9	-0.6	-0.3	0.0	-0.5
(34,42)	0.4	-0.9	-0.8	-0.4	0.3	-0.4
(43,51)	-0.8	-0.9	-0.4	-0.3	0.2	-0.5
(52,60)	-0.9	-0.6	-0.5	-0.1	0.2	-0.5
All	-0.9	-1.1	-0.6	-0.3	0.1	-0.6
			Excess kurtosis			
(15,24)	19.1	13.4	1.8	0.7	0.3	8.0
(25,33)	8.4	2.7	1.0	0.4	0.1	2.8
(34,42)	13.3	5.0	3.2	1.2	0.8	5.1
(43,51)	12.8	5.7	1.2	0.7	0.1	4.6
(52,60)	12.9	4.1	1.9	0.1	0.1	4.5
All	13.5	6.7	1.8	0.6	0.3	5.1
			First lag autocorrelations			
(15,24)	0.12	0.03	0.00	-0.03	-0.03	0.02
(25,33)	0.13	0.02	-0.01	-0.03	0.03	0.02
(34,42)	0.07	-0.01	0.01	0.01	0.04	0.02
(43,51)	0.06	0.00	-0.10	-0.01	0.00	-0.02
(52,60)	0.07	0.00	0.01	-0.04	-0.01	0.01
All	0.10	0.01	-0.01	-0.02	0.00	0.02
			First lag autocorrelations of the squares			
(15,24)	0.13	0.07	0.09	0.10	0.01	0.09
(25,33)	0.27	0.13	0.13	0.07	0.06	0.14
(34,42)	0.11	0.06	0.05	0.03	0.06	0.06
(43,51)	0.08	0.14	0.12	0.10	-0.11	0.10
(52,60)	0.20	0.06	0.09	0.12	0.07	0.11
All	0.17	0.09	0.10	0.08	0.02	0.10

correlations across all contracts.

The first two panels of Table 2.9 consider the sample skewness and excess kurtosis of the pricing residuals. The estimates based on predicted residuals can be used to test the normality hypothesis of the measurement errors. Normality does not seem to be completely at odd with the data, with the exception of out-of-the-money options, for which the symptoms of negative skewness and leptokurtosis are clear.

Finally, the last two panels of Table 2.9 report the first order sample autocorrelations between the pricing errors and their values lagged one period, as well as autocorrelations between consecutive squared pricing residuals. The results for predictive residuals highlight some dynamic misspecification for deep out-of-the-money options; the evidence is fairly coherent to the IID hypothesis in the other cases. Autocorrelations are much higher for residuals filtered conditionally on augmented information sets (not reported), but our results are again coherent with those reported elsewhere – see e.g. the two columns under the heading "Autocorrelations" in table 2 of Bates (2000).

To summarize: the analysis of the generalized residuals suggests that the stock price model is probably misspecified, and should allow at least for a more flexible form of heteroskedasticity. The filtered predicted option pricing residuals are not incompatible with the IID assumption, but point to the need to extend the model in order to explain the very high contemporaneous correlations reported in table 2.8. Finally, dramatic improvements in predicting option prices can be obtained by pricing options relative to other options at the same date. Using the results in tables 2.6 and 2.7, it can be seen that conditioning on all the other options at the same date reduces the RMSE by a factor ranging from 9% to 58%, depending on the class of moneyness and time to maturity, with respect to prices computed conditioning on just one option.

2.6 Conclusions

In this paper, we consider joint estimation of objective and risk-neutral parameters for SV option pricing models using both stock and option prices. A common strategy simplifies the task by limiting the analysis to just one option per date. We first discuss its drawbacks on the basis of model interpretation, estimation results and pricing exercises. We then turn the attention to a more flexible approach, that successfully exploits the wealth of information contained in large heterogeneous panels of options, and we apply it to actual S&P 500 index and index call options data.

Our approach has two crucial features. First, we break the stochastic singularity between contemporaneous option prices by assuming that every observation is affected by measurement error. We deem this assumption much more appealing that the alternative one, in which at each date one specific option is observed without measurement error. The price to pay for this increased flexibility is that the evaluation of the likelihood function poses some non trivial numerical challenges, but we successfully overcome them using a MC-IS strategy, combined with a Particle Filter algorithm. Second, we approximate the theoretical model-implied option prices using a highly flexible parametric model, which allows us to compute very quickly and accurately a huge number of implied volatilities.

The results we obtain suggests that the model is misspecified, but that some significant improvements could probably be obtained by extending it in the direction of including jumps or regime switching in the volatility dynamics. Other extensions can be envisioned, but we leave them to future research.

Bibliography

Aït-Sahalia, Y. (2008), "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36, 906-937.

Aït-Sahalia, Y. and Kimmel, R. (2007), "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413-52.

Aït-Sahalia, Y. and Kimmel, R. (2010), "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions," *Journal of Financial Economics*, 98, 113-144.

Andersen, T. G., Benzoni, L. and Lund, J. (2001), "Towards an Empirical Foundation of Continuous-Time Equity Return Models," *The Journal of Finance*, 57, 1239-1284.

Bakshi, G., Cao, C. and Chen, Z. (1997), "Empirical Performance of Alternative Option Pricing Models," *The Journal of Finance*, 52(5), 2003-2049.

Bates, D. S. (2000), "Post-'87 Crash Fears in the S & P 500 Futures Option Market," *Journal of Econometrics*, 94, 181-238.

Bates, D. S. (2003), "Empirical Option Pricing: a Retrospection," *Journal of Econometrics*, 116, 387-404.

Black, F. and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-659.

Brandt, M.W. and He, P. (2005), "Simulated Likelihood Estimation of Affine Term Structure Models from Panel Data," working paper Wharton School of Business, University of Pennsylvania.

Broadie, M., Chernov, M. and Johannes, M. (2007), "Model Specification and Risk Premia: Evidence from Futures Options," *The Journal of Finance*, 62(3), 1453-1490.

Cheng, P. and Scaillet, O. (2007), "Linear Quadratic Jump-Diffusion Modeling," *Mathematical Finance*, 17(4), 575–598.

Chernov, M., Gallant, R. A., Ghysels, E. and Tauchen, G. (2003), "Alternative Models of Stock Price Dynamics," *Journal of Econometrics*, 116, 225-257.

Christoffersen, P. Jacobs, K. and Mimouni, K. (2010), "Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns and Option Prices," *The Review of Financial Studies*, 23(8), 3141-3189.

Conley, T. G., Hansen, L. P., Luttmer, E. G. J. and Scheinkman, J. A. (1997), "Short-Term Interest Rates as Subordinated Diffusions," *The Review of Financial Studies*, 10, 525-577.

De Jong, F. (2000), "Time Series and Cross-Section Information in Affine Term-Structure Models," *Journal of Business & Economic Statistics*, 18, 300-14.

Duffee, G. (2002). "Term Premia and Interest Rate Forecasts in Affine Models," *The Journal of Finance*, 57 (1), 405–443,

Duffie, D., Pan, J. and Singleton, K. (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343-76.

Durham, G.B. and Gallant, R.A. (2002), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes," *Journal of Business & Economic Statistics*, 20(3), 297-316.

Durham, G.B. (2006), “Monte Carlo Methods for Estimating, Smoothing, and Filtering One and Two-Factor Stochastic Volatility Models,” *Journal of Econometrics*, 133, 273-305.

Durham, G.B. (2007), “SV Mixture Models with Application to S&P 500 Index Returns,” *Journal of Financial Economics* 85, 822-856.

Durham, G.B. (2010), “Risk-Neutral Modelling with Affine and Non-Affine Models,” working paper Leeds School of Business, University of Colorado at Boulder.

Gelman A., Carlin J., Stern H. and Rubin D. (1995), *Bayesian Data Analysis*, Chapman&Hall, London.

Geweke, J. (1989), “Bayesian Inference in Econometric Models Using Monte Carlo Integration,” *Econometrica*, 24, 1037-1399.

Huber, P., Scaillet, O. and Victoria-Feser, M.-P. (2009), “Assessing Multivariate Predictors of Financial Market Movements: A Latent Factor Framework for Ordinal Data,” *The Annals of Applied Statistics*, 3(1), 249-271.

Jiang, G. and Tian, Y. (2005), “The Model-Free Implied Volatility and Its Information Content,” *The Review of Financial Studies*, 18, 1305-1342.

Jiang, G. and Tian, Y. (2007), “Extracting Model-Free Volatility from Option Prices,” *Journal of Derivatives*, 14(3), 35-60.

Jones, C. S. (2003), “The Dynamics of Stochastic Volatility: Evidence From Underlying and Options Markets,” *Journal of Econometrics*, 116, 181-224.

Koopman, S. J., Shephard, N. and Creal, D. (2009), “Testing the Assumptions Behind Importance Sampling,” *Journal of Econometrics*, 149, 2-11.

Lee, R. W. (2001), “Implied and Local Volatilities Under Stochastic Volatility,” *International Journal of Theoretical and Applied Finance*,” 4(1), 45-89.

Lewis, A.S. (2000), *Option Valuation under Stochastic Volatility*, Finance Press.

Medvedev, A. and Scaillet, O. (2007), "Approximation and Calibration of Short Term Implied Volatilities Under Jump-Diffusion Stochastic Volatility," *The Review of Financial Studies*, 20(2), 427-459.

Merton, R. C. (1976), "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3, 125-44.

Monahan, J. F. (1993), "Testing the Behavior of Importance Sampling Weights," *Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface*," 112-117.

Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3-50.

Pastorello, S., Patilea, V. and Renault, E., (2003) "Iterative and Recursive Estimation in Structural Nonadaptive Models," *Journal of Business & Economic Statistics*, 21(4), 449-509.

Pitt, M. K. (2002), "Smooth Particle Filters for Likelihood Evaluation and Maximization," working paper, University of Warwick.

Rao, C. R. (1952), *Advanced Statistical Methods in Biometric Research*, Wiley, New York.

Rubin, D. B. (1988), "Using the SIR Algorithm to Simulate Posterior Distributions," in J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds), *Bayesian Statistics*, 3, 395-402, Oxford University Press.

Sircar, K. R. and G. Papanicolau (1999), "Stochastic Volatility, Smile and Asymptotics," *Applied Mathematical Finance*, 6, 107-45.

Tauchen, G. (2002), "Comment on Durham and Gallant (2002)," *Journal of Business & Economic Statistics*, 20(3), 331-332.

Willard, G. A. (1997), "Calculating Prices and Sensitivities for Path-Independent Derivative Securities in Multifactor Models," *Journal of Derivatives*, 5(1), 45-61.

2.A The data

The sample contains data on the spot price of the S&P500 index and daily call prices on the index. It consists of 2,517 daily observations from January 4, 1996 to December 30, 2005. For each date, we extracted from the population of all exchanged call options on the index (179,176 over the whole 10 years interval) those with discounted moneyness $|X_{it}| \leq 5\%$, time to maturity $15 \leq \tau_{it} \leq 60$ days, and transaction volume of at least 5 contracts, leaving us with a total of 40,211 observed options. The constraints were imposed to exclude from the sample illiquid or seldom traded contracts. The number of call options observed at each date varies from 3 to 44, with an average of almost 16. It also appears that the cross-sectional dimension steadily grew over the last four observation years.

Table 2.10: Sample frequencies of option observations computed according to different sample selection rules by discounted moneyness and maturity.

Time to	Discounted Moneyness X					
Maturity τ	(-5%,-3%)	(-3%,-1%)	(-1%,1%)	(1%,3%)	(3%,5%)	All
(15,24)	2,384	2,647	2,693	1,965	1,219	10,908
(25,33)	2,205	2,414	2,540	1,878	1,213	10,250
(34,42)	1,472	1,702	1,872	1,249	766	7,061
(43,51)	1,200	1,422	1,588	966	534	5,710
(52,60)	1,406	1,564	1,760	1,018	534	6,282
All	8,667	9,749	1,0453	7,076	4,266	40,211

Table 2.10 reports the sample frequencies of options observations by discounted moneyness and maturity. This table highlights that the sample contains a clear prevalence of short maturity options, and an even clearer majority of at- and out-of-the-money options.

2.B The MC-IS approximation of the transition density

To illustrate the IS approach we use, we follow Durham (2010, sect. 3.2).

Let $t-1=\tau_0<\tau_1<...<\tau_M=t$ be a partition of the interval $[t-1,t]$, where for simplicity we assume that for all m, $\tau_m-\tau_{m-1}=\Delta/M=\delta$, where Δ is the length of the interval between observations at dates $t-1$ and t. Let us for simplicity denote $P_m=P_{\tau_m}$ and $V_m=V_{\tau_m}$, $m=0,1,\ldots,M$, and $\boldsymbol{P}=(P_1,\ldots,P_{M-1})'$ and $\boldsymbol{V}=(V_1,\ldots,V_{M-1})'$. By the Chapman-Kolmogorov property and the Markov nature of the bivariate diffusion:

$$f(P_t,V_t|P_{t-1},V_{t-1})=\int\prod_{m=1}^{M}f(P_m,V_m|P_{m-1},V_{m-1})\,d\boldsymbol{P}d\boldsymbol{V}.$$

This integral can be approximated using

$$f^{(M)}(P_t,V_t|P_{t-1},V_{t-1})=\int\prod_{m=1}^{M}f^a(P_m,V_m|P_{m-1},V_{m-1})\,d\boldsymbol{P}d\boldsymbol{V}.$$

where f^a is the transition density implied by an approximate discretization scheme of (2.5) - (2.4); in this paper, we use the first order Euler scheme, which implies that f^a is bivariate Gaussian. Apart from its simplicity, this choice also allows to analytically compute the integral with respect to $\boldsymbol{P}$, leaving us with:

$$f^{(M)}(P_t,V_t|P_{t-1},V_{t-1})=\int f^a(P_M|P_0,V_M,\boldsymbol{V},V_0)\,f^a(V_M,\boldsymbol{V}|V_0)\,d\,\boldsymbol{V} \tag{B-1}$$

where $f^a(V_M,\boldsymbol{V}|V_0)$ is the product of M Gaussian densities:

$$f^a(V_M,\boldsymbol{V}|V_0)=\prod_{m=1}^{M}\phi[V_m;V_{m-1}+\mu_V(V_{m-1})\,\delta,\sigma_V^2(V_{m-1})\,\delta]$$

where ϕ is the Gaussian density, and $f^a(P_M|P_0,V_M,\boldsymbol{V},V_0)$ is the conditional distribution of P_M given $P_0,V_M,\boldsymbol{V}$ and V_0 implied by the Euler discretization. Under our assumptions about the jump process, this density is a mixture of Gaussian pdfs, with the number of jumps being the mixing variable:

$$f^a(P_M|P_0,V_M,\boldsymbol{V},V_0)=\sum_{n=0}^{\infty}\frac{e^{-\lambda_J\Delta}(\lambda_J\Delta)^n}{n!}\,\phi[P_M;P_0+k(V_M,\boldsymbol{V})+n\mu_J,s^2(\boldsymbol{V})+n\sigma_J^2]$$

where:

$$k(V_M, \boldsymbol{V}) = \sum_{m=1}^{M} \mu_{P|V}(V_m, V_{m-1})\delta, \quad s^2(\boldsymbol{V}) = \sum_{m=1}^{M} \sigma^2_{P|V}(V_{m-1})\delta,$$

with:

$$\mu_{P|V}(V_m, V_{m-1})\delta = \mu_P(V_{m-1})\delta + \frac{\rho\sigma_P(V_{m-1})}{\sigma_V(V_{m-1})} \left[V_m - V_{m-1} - \mu_V(V_{m-1})\delta\right],$$

$$\sigma^2_{P|V}(V_{m-1})\delta = \sigma^2_P(V_{m-1})\delta\,(1-\rho^2).$$

The $(M-1)$-dimensional integral in (B-1) can be evaluated using a IS approach. Let $q(\boldsymbol{V})$ be a pdf on $\mathbb{R}^{M-1}$, and rewrite (B-1) as follows:

$$f^{(M)}(P_t, V_t|P_{t-1}, V_{t-1}) = \int f^a(P_M|P_0, V_M, \boldsymbol{V}, V_0)\, \frac{f^a(V_M, \boldsymbol{V}|V_0)}{q(\boldsymbol{V})}\, q(\boldsymbol{V})d\boldsymbol{V}. \qquad \text{(B-2)}$$

Let $\tilde{\boldsymbol{V}}^l = (\tilde{V}^l_1, ..., \tilde{V}^l_{M-1})$, $l = 1, ..., L$, be independent draws from q. The IS estimate of (B-2) is then given by

$$\tilde{f}^{(M,L)}(P_t, V_t|P_{t-1}, V_{t-1}) = \frac{1}{L}\sum_{l=1}^{L} f^a(P_M|P_0, V_M, \tilde{\boldsymbol{V}}^l, V_0)\, \frac{f^a(V_M, \tilde{\boldsymbol{V}}^l|V_0)}{q(\tilde{\boldsymbol{V}}^l)}. \qquad \text{(B-3)}$$

Durham and Gallant (2002) showed that to reduce the bias in (B-3) it is important to transform the original diffusion into another one characterized by a constant diffusion coefficient. In our framework, it is sufficient to consider the Lamperti transform of V_t, which is defined as:

$$Y_t = \int^{V_t} \frac{1}{\sigma_V(u)}\, du$$

where the lower bound is irrelevant. By Itô's Lemma:

$$dY_t = \left[\frac{\mu_V(V_t)}{\sigma_V(V_t)} - \frac{1}{2}\frac{\partial\sigma_V(V_t)}{\partial V_t}\right] dt + dW_{Vt}.$$

The variance of the simulation noise can be reduced by carefully choosing q. In this respect, a very efficient sampling strategy, labelled Modified Brownian Bridge (MBB) by Durham and Gallant (2002), suggests to recursively draw $\tilde{V}_m$ for $m = 1, ..., M-1$ from a Gaussian density based on the Euler discretization of the process and conditional to V_M and V_{m-1}. With a minor approximation, this density is Gaussian with mean equal to $V_{m-1}+(V_M-V_{m-1})/(M-m+1)$, and variance given by $[(M-m)/(M-m+1)]\,\sigma_V^2(V_{m-1})\,\delta$. The product of these $M-1$ Gaussian densities defines the auxiliary density which is used as the denominator in (B-3), and as the distribution from which the simulated $\tilde{\boldsymbol{V}}^l$, $l = 1, ..., L$ are drawn. Sampling from it is extremely fast and can be combined with other variance reduction techniques. On the basis of the analysis in Durham (2010), we implement the MBB approach using $M = 8$ subintervals, and $L = 256$ simulated volatility trajectories. A comparison across parameter estimates of the unavoidable sampling noise with the simulation noise confirms the adequacy of these settings.

2.C Numerical evaluation of option prices in non-affine jump-diffusion models

Apart a few special cases, theoretical option prices are not known in closed-form in a general jump-diffusion SV model, and need to be evaluated numerically. In the absence of jumps, a particularly efficient pricing strategy was advanced by Willard (1997), based on conditional Monte Carlo technique combined with quasi random sequences. This strategy is based on the observation that, in a pure SV model, the price of a call can be computed by first conditioning on the trajectory of the volatility Brownian motion during the life of the option, and then numerically evaluating the expectation with respect to the distribution of the trajectory. The resulting pricing formula $C_{SV}(S_t, V_t)$ is then given by:

$$C_{SV}(S_t, V_t) = \mathrm{E}_{W^{\mathbb{Q}}_{V[t,t+\tau]}} \left\{ C_{BS}\left[S_t \xi_{[t,t+\tau]}, (1-\rho^2)\tilde{\sigma}^2_{[t,t+\tau]} \right] \right\},$$

where $C_{BS}(S,\sigma^2)$ is the Black and Scholes formula for price S and volatility σ^2 (and where for simplicity we neglect the remaining arguments), and:

$$\begin{aligned}
\xi_{[t,t+\tau]} &= \exp\left[-\frac{\rho^2}{2}\int_t^{t+\tau}\sigma_S^2(V_u)\,du+\rho\int_t^{t+\tau}\sigma_S(V_u)\,dW_{Vu}^{\mathbb{Q}}\right],\\
\tilde{\sigma}^2_{[t,t+\tau]} &= \frac{1}{\tau}\int_t^{t+\tau}\sigma_S^2(V_u)\,du.
\end{aligned}$$

Merton (1976) used a similar argument to price options with (i) Poisson jumps in the S_t process with intensity $\lambda_J^{\mathbb{Q}}$ and size independent from the Brownian processes, and (ii) relative jump size equal to $e^{J^{\mathbb{Q}}}-1$, where $J^{\mathbb{Q}}\sim\mathcal{N}(\mu_J^{\mathbb{Q}},\sigma_J^{\mathbb{Q}^2})$. In our case, it is possible to proceed by conditioning on n of jumps during the life τ of the option, compute the price using the Willard (1997) approach with modified arguments, and then compute the expectation with respect to a Poisson distribution with parameter $\lambda_J^{\mathbb{Q}}\tau$:

$$C_{SVJ}=\sum_{n=0}^{\infty}\frac{e^{-\lambda_J^{\mathbb{Q}}\tau}(\lambda_J^{\mathbb{Q}}\tau)^n}{n!}\,\mathrm{E}_{W^{\mathbb{Q}}_{V[t,t+\tau]}}\left\{C_{BS}\left[S_t\xi_{[t,t+\tau]}c_n,(1-\rho^2)\tilde{\sigma}^2_{[t,t+\tau]}+\frac{n\sigma_J^{\mathbb{Q}^2}}{\tau}\right]\right\}, \quad \text{(C-1)}$$

where:

$$c_n=\exp\left[n\nu_J^{\mathbb{Q}}-\lambda_J^{\mathbb{Q}}\tau(e^{\nu_J^{\mathbb{Q}}}-1)\right].$$

A truncated version of (C-1) provides a viable strategy to compute the option price.

Chapter 3

Informed and Uninformed Traders at Work: Evidence from the French Market

Fabrizio Ferriani

JEL codes: C10, G10, G14

Keywords: Market microstructure, Euronext Paris, Informational asymmetries, Price impact, Trader identities

3.1 Introduction

In recent years, there has been an increasing attention given to the field of market microstructure as evidenced by the significant amount of contributions to theoretical and empirical literature. Several factors facilitated this process, including the availability of high-quality data sets with transaction-level information, the diffusion of new markets, and the interest for the trading strategies of market participants. This paper adds to the existing literature by investigating how different categories of traders affect market transactions in a high-frequency context. In particular, we distinguish between informed and uninformed traders, and we analyze their impacts on market prices. This topic is of primary interest for the empirical research devoted to informational issues in financial markets, which represents a significant concern for traders and market regulators worldwide.

Theoretically speaking, we trace back to Kyle (1985) for a first relevant study of the traders' behaviour in an order-driven market. Kyle (1985) derives a model where a risk-neutral market maker operates with an insider and a group of liquidity traders. The market maker observes the whole order flow and he sets an equilibrium price that internalizes the presence of informed traders. An extension of this setting is found in Admati and Pfleiderer (1988) who propose a similar framework, but they distinguish between two groups of liquidity traders. Conversely, Glosten and Milgrom (1985) introduce a sequential model where information asymmetries take place in a quote-driven protocol. In their model, the risk of trading with informed agents forces the dealers to widen the distance between the bid and the ask quote. Indeed, the spread operates as a protection for the adverse selection cost and its width is proportional to the fraction of informed traders in the market. This model is extended by Easley and O'Hara (1987) to take into account the order size. The authors emphasize the information effect driven by the trading volume, with insiders submitting large transaction volumes to exploit their private signal.

In the field of empirical research, Glosten and Harris (1988) study the components of the bid-ask spread on a sample of stocks from the NYSE. They find the information asymmetries to significantly affect the bid-ask spread width. Easley et al. (1996) and Easley et al. (1997) introduce the concept of PIN, which offers a measure of the informed traders' presence in the market. In their model, only some agents receive an information signal; a bayesian market maker observes the sequence and the frequency of trades, then he updates the quotes according to his belief on the type of information event. Their model has been very successful and it has been widely tested in different markets and extended in several ways, e.g. Easley et al. (2002) or Easley et al. (2008). The activity of informed traders is also investigated by Barclay and Warner (1993), Chakravarty (2001) and Alexander and Peterson (2007) in the field of the stealth trading literature. These papers examine the relationship between order size and information asymmetries and they find that informed traders tend to submit medium-sized orders to disguise their identity. In fact, and because of volume informativeness, large orders are easily interpreted as an attempt to quickly maximize the profits of a private information signal; on the contrary, small volumes would be inconvenient because of transaction and time costs. Finally, Foucault et al. (2007) examine the effects of a switch from a fully-disclosed market to a regime with hidden identities, in which the informed traders are not explicitly detectable. They consider the market reform that has interested Euronext Paris in 2001, and they find that the switch to the new market setting reduces the average spread, increases the liquidity of the market, and diminishes the informativeness of the limit order book (LOB) variables.

In this paper, we explore informational issues at Euronext Paris and we base our empirical research on Hausman et al. (1992), who propose an ordered probit scheme to analyse the transaction prices. We extend their approach to exploit the information on traders' identities at the transaction level. We find that informed traders significantly affect the stock price when trading with uninformed agents. This result may seem surprising after the introduction of the anonymous regime in 2001. Therefore, to explain our findings, we dedicate the second part of this research to analyse the informativeness of a wide set of LOB variables, and we find that most of them are highly and unambiguously significant to infer the identity of traders.

The paper is organized as follows. Section 3.2 presents the model employed in the estimation, while Section 3.3 describes the data used for the empirical analysis. Section 3.4 details the variables included in the empirical specification and how the trader effect is incorporated in the model. Section 3.5 examines the parameter estimates, Section 3.6 focuses on postestimation issues and marginal effects, while Section 3.7 presents the robustness tests. Section 3.8 discusses the informativeness of LOB variables for the inference of traders' identities. Lastly, Section 3.9 concludes.

3.2 The Model

As anticipated in the Introduction, we generalize Hausman et al. (1992), hereafter HLM (1992), to evaluate the impact of trader categories on the transaction prices. In the following lines, we briefly review the original approach, but we refer the reader to HLM (1992) for a more exhaustive exposition. We consider a sequence of transaction prices $P_{t_0}, P_{t_1}, ..., P_{t_n}$, observed at times $t_0, t_1, ..., t_n$, where each t_i corresponds to the effective transaction time, without reference to a fixed sampling frequency. We define the variable tick D_{t_k} as the difference between two consecutive prices multiplied by 100, i.e. $D_{t_k} = (P_{t_k} - P_{t_{k-1}}) * 100$[1]. For our sampling period, the tick size at Euronext Paris equals to 0.01 Euro; hence, D_k provides the price change expressed in Euro cents. In an ordered probit model, D_k can be thought of as the observed realization of a latent continuous random variable:

$$D_k^* = \mathbf{X}_k'\beta + \epsilon_k \tag{3.1}$$

where $\mathbf{X}_k$ includes the variables that influence the mean of D_k^*, while ϵ_k is a Gaussian noise with zero mean and variance equal to $\sigma_k^2 = \mathbf{W}_k'\theta$, with $\mathbf{W}_k$ including the variables that affect the variance. The two vectors β and θ collect the parameters associated with the mean and the variance regressors, respectively. The relationship linking the observed variable D_k with its latent counterpart D_k^* is governed by the subsequent interval classification

$$D_k = \begin{cases} d_1, & \text{if } D_k^* \in A_1 =]-\infty; \alpha_1], \\ d_2, & \text{if } D_k^* \in A_2 =]\alpha_1; \alpha_2], \\ \vdots & \vdots \\ d_m, & \text{if } D_k^* \in A_m =]\alpha_{m-1}; \infty[\end{cases}$$

where $\alpha_1 < \alpha_2 < ... < \alpha_{m-1}$ represent non-overlapping cut points dividing the whole data range of D_k^* into m distinct intervals A_j, $j = 1, ..., m$, while d_j identifies the possible

[1] In the following, we will use only k instead of t_k to simplify the notation.

outcomes of the observed price change D_k. Under the assumption of conditional independence and Gaussianity of the error distribution, the model may be easily estimated by ML. We define $\gamma' = [\beta', \theta', \alpha_1, ..., \alpha_{m-1}]$ as the vector of all the parameters included in the model together with D_k^* thresholds. The likelihood function to be maximized is described by

$$\sum_{k=1}^{n} \left\{ Y_{1k} \cdot \log \Phi\left(\frac{\alpha_1 - \mathbf{X}_k' \beta}{\sigma_k}\right) + \sum_{i=2}^{m-1} Y_{ik} \cdot \log \left[\Phi\left(\frac{\alpha_i - \mathbf{X}_k' \beta}{\sigma_k}\right) - \Phi\left(\frac{\alpha_{i-1} - \mathbf{X}_k' \beta}{\sigma_k}\right)\right] \right.$$
$$\left. + Y_{mk} \cdot \log \left[1 - \Phi\left(\frac{\alpha_{m-1} - \mathbf{X}_k' \beta}{\sigma_k}\right)\right] \right\} \tag{3.2}$$

where Y_{ik} is an indicator variable equal to one if D_k belongs to the $i-th$ interval, and Φ is the standard normal cumulative distribution function.

The ordered probit model represents the reference estimation tool adopted in this paper, though Section 3.5 also considers different specifications to verify the robustness of results. As an alternative choice among the feasible estimation frameworks, we could have considered the interval regression model, which exhibits a remarkable similarity with the ordered probit. Nevertheless, this correspondence is only apparent and some substantial differences exist between the two approaches. First, the dependent variable in the ordered probit model does not have a quantitative connotation, and the interpretation of the marginal effects does not coincide in the two cases. In fact, the marginal effect in the ordered probit is nonlinearly related to the whole set of regressors; conversely, the estimated coefficients in the interval regression model directly express the marginal contribution of each variable under the *ceteris paribus* condition. This aspect is also reflected in the diagnostics of interest between the two models: interval regression is more suitable to study the effect of the explanatory variables on the conditional mean, while the ordered probit is more appropriate to examine the impact on the conditional probability distribution of D_k^*. Second, a relevant distinction between the two approaches concerns the dimension of the γ' vector. In the interval regression, the cut points do not enter in the set of parameters to be estimated, but they are pre-determined by the researcher or by the sampling procedure. Indeed, this last point is far from being a minor issue. Excluding the cut points from γ' is a practicable choice only if the range of variation of D_k^* is defined without ambiguity or arbitrariness. Moreover, as emphasized by HLM (1992), the estimation of the cut points together with the other model parameters allows to fully describe the (nonlinear) relationship between the observed realizations and the latent variable. All these motivate us to adopt the ordered probit as the basic estimation tool.

3.3 The Data

3.3.1 Descriptive analysis

The data set for the empirical analysis is provided by Eurofidai, and it consists of all the transactions registered for the stocks of the CAC 40 index at Euronext Paris from 3 February 2008 to 31 March 2008. We remove the records relative to the opening and the closing auctions, and we only focus on the transactions executed during the continuous trading session, i.e. from 9.00 a.m. to 5.30 p.m..

In Table 3.1, we report some descriptive statistics for the stocks included in the CAC 40 index. The first column displays the market capitalization, while the other four columns exhibit the total amount of transactions, the daily average number of transactions, the average transaction volume, and the average price, respectively. In Figure 3.1 and 3.2, we also provide some descriptive plots for Bouygues and BNP; for the sake of brevity, we limit the exposition to these two stocks, as the rest of the sample presents similar patterns.

Table 3.1: Descriptive statistics for CAC40 index stocks. Stocks are ordered according to market capitalization on 3 February 2008. The market capitalization is expressed in Euro millions.

	Market Cap	Num. of Trans.	Avg. Trans.	Avg. Vol.	Avg. Pr.
Capgemini	5,245	258,998	6,475	306	35.89
Technip	5,289	181,423	4,536	206	51.02
AF-KLM	5,686	206,466	5,162	469	17.56
STM	6,135	121,228	3,031	2,130	7.66
Lagardere	6,352	122,307	3,058	211	49.53
Vallourec	8,155	330,792	8,270	91	139.81
Alcatel	8,387	205,301	5,133	3,524	3.90
Essilor	8,744	134,672	3,453	232	39.15
Michelin	9,521	296,959	7,424	218	63.38
Accor	10,635	236,324	5,908	265	48.00
Peugeot	11,505	278,311	6,958	308	49.41
Ppr	12,020	175,130	4,378	153	90.85
Eads	12,218	270,256	6,756	704	16.67
Unibail	13,329	189,694	4,742	106	162.10
Bouygues	13,983	274,566	6,864	269	45.72
Pernod	14,301	186,605	4,665	172	69.19
Lafarge	19,009	281,096	7,027	110	151.13
Saint Gobain	19,328	379,089	9,477	274	51.15
Alstom	19,367	324,377	8,109	101	138.57
Renault	19,974	404,396	10,110	241	69.50
Schneider	20,093	346,825	8,671	186	77.65
Veolia	20,786	383,532	9,588	289	51.30
Vinci	22,299	310,050	7,751	276	45.03
Air Liquide	22,795	261,579	6,539	134	93.25
Vivendi	28,827	389,910	9,748	660	25.82
Danone	29,047	355,751	8,894	317	53.67
Crédit Agricole	32,727	399,358	9,978	777	18.45
Carrefour	34,448	308,801	7,720	361	47.13
Lvmh	34,540	310,584	7,764	228	68.57
Société Generale	36,174	803,620	20,090	407	70.80
Gdf	37,623	219,327	5,483	269	37.44
Axa	47,376	537,978	13,449	1,035	21.80
L'Oréal	49,131	279,003	6,975	194	80.39
Suez	54,333	422,244	10,556	435	41.48
France Télécom	55,568	492,229	12,306	1,034	22.47
BNP	57,864	690,621	17,266	353	60.57
Sanofi-Aventis	64,908	440,472	11,012	439	49.91
Arcelor	75,165	366,871	9,172	568	49.13
Edf	100,419	411,595	10,290	235	63.10
Total	112,685	591,689	14,792	545	48.79

Figure 3.1: Frequency distribution of D_k (price change) and Δt_k (trade durations). The vertical scale has been reduced to make the figure more intelligible; the bold number indicates either the frequency of $D_k = 0$ or the one of $\Delta t_k = 0$.

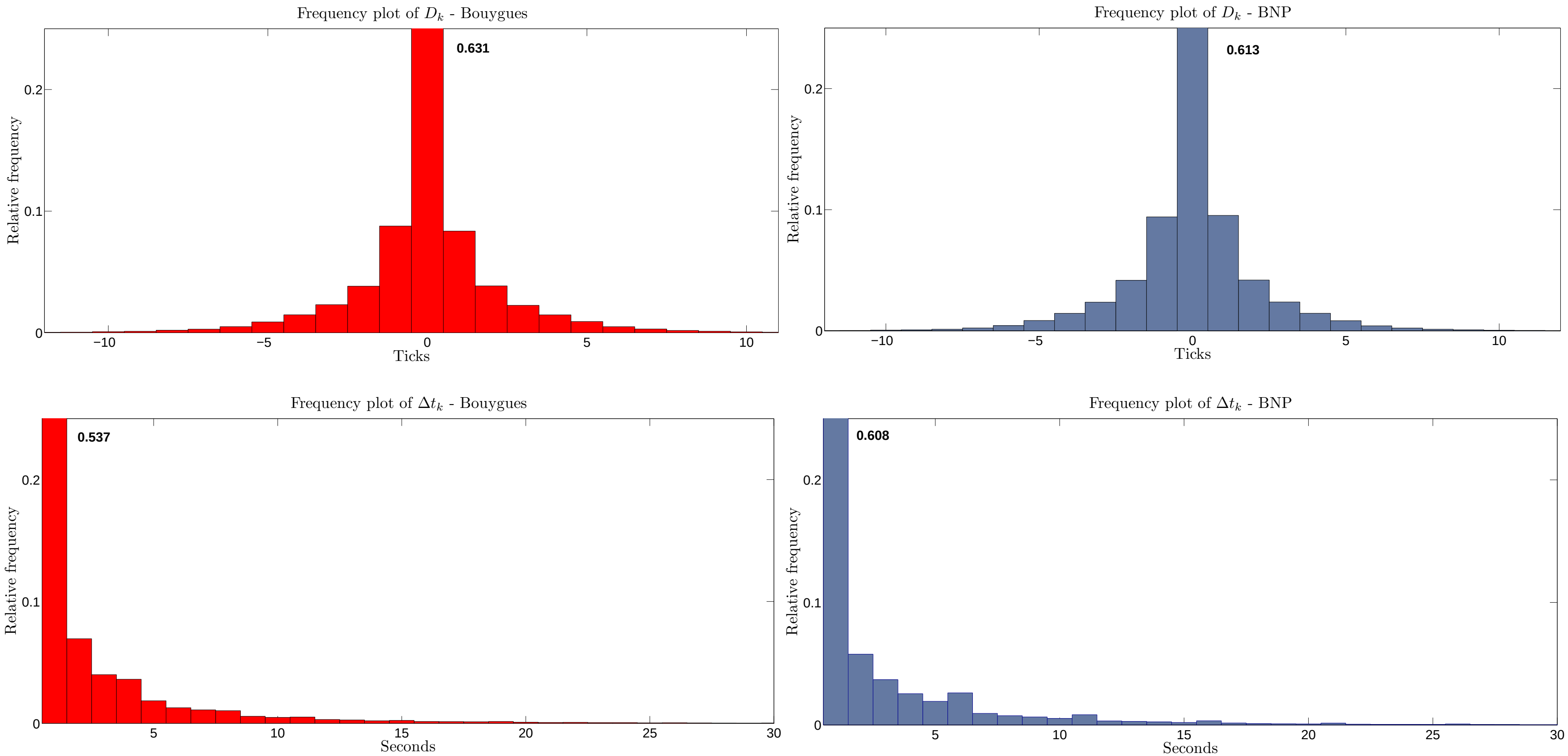
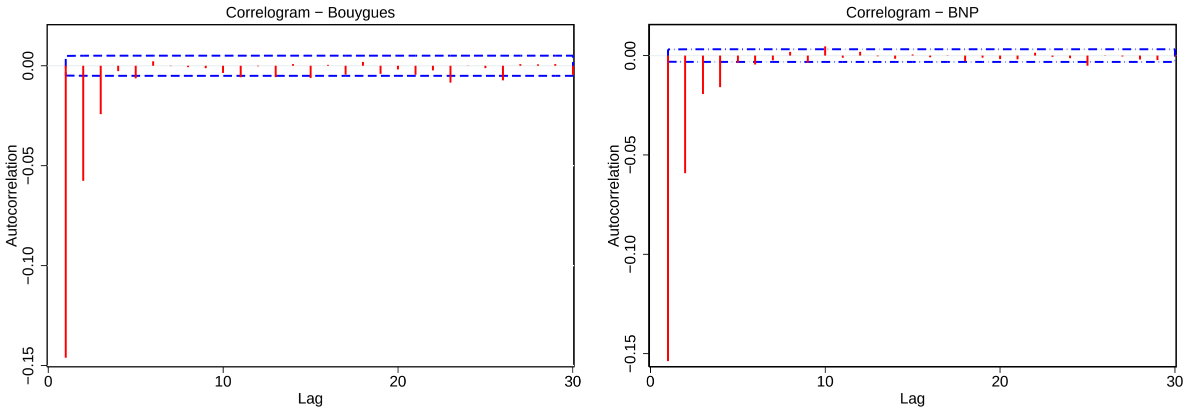


Figure 3.2: 30-lag correlogram of D_k. The straight lines represent confidence interval at 99%.



The upper panel of Figure 3.1 displays the frequency distribution of the price variation D_k, while the lower panel shows the frequency distribution of trade durations (Δt_k) expressed in seconds. The upper panel of Figure 3.1 shows a noticeable concentration of values corresponding to $D_k = 0$, similar to previous researches in this field, such as Liesenfeld et al. (2006). This stylized fact is related to the specificity of high-frequency financial data, which report transactions occurring within very short time intervals. This feature clearly rules out the possibility of large and frequent jumps in prices, especially for highly liquid stocks where the depth of the LOB assures the order execution within limited price skips. Some stocks exhibit certain peculiarities, wherein the distributions are either slightly skewed or they are characterized by thicker tails and higher dispersion.

Then, the lower panel of Figure 3.1 presents the frequency distribution of trade durations. High-frequency data usually display very short durations between consecutive transactions[2], which is particularly true for liquid stocks as the ones of the CAC 40 index. The two plots in the lower panel exhibit a remarkable left-skewness, with a high concentration of transactions occurring simultaneously or within very small time intervals. Finally, Figure 3.2 provides the 30-lag correlogram of D_k that clearly shows a

[2]Transaction time is expressed as 'hh:mm:ss'. For simultaneous observations at the level of time seconds, the data set provides a code to chronologically sort the transactions, independently of the time of execution.

negative autocorrelation pattern, at least for the first lags. This negative autocorrelation is a standard feature that occurs in the whole sample, and it has been normally justified through the fluctuations between bid and ask quotes (Roll, 1984).

3.3.2 The categories of traders

For each observation, the data set displays a code to identify the different categories of traders:

- '1' is the code that is attributed to transactions executed on behalf of retail investors;
- '2' is the code that refers to transactions executed by operators authorized to trade in the Paris Bourse. This code includes banks or other financial intermediaries, called 'Sociétés de Bourse';
- '6' is the code that classify transactions executed by 'fournisseurs de la liquidité', i.e. agents with liquidity duties;
- '7' is the code that categorizes the transactions executed by another kind of financial intermediaries called 'Filiales de la Société de Bourse'. They are financial institutions similar to the traders classified with code '2'.

Generally speaking, all the transactions are executed by authorized operators, i.e. stock members, but only trades coded with '1' are executed on behalf of retail investors. Observations classified with '2' or '7' refer to transactions executed by financial intermediaries in their own interest. The code is available for both sides of the market, such that the buyer and the seller can be of the same kind of operator or they can belong to different categories. Since this data set involves highly liquid stocks, there is usually no need for a liquidity provider, and the number of transactions registered with code '6' is generally absent or extremely limited. All the transactions registered with '1' are attributed to retail or uninformed traders, while all the transactions registered with '2' and '7' are placed in the category of institutional or informed traders. This is a plausible distinction that follows the one proposed by Chakravarty (2001) in his analysis of stealth trading at the NYSE. For our analysis, we recall that Euronext LOB has been completely anonymous starting from 2001, and the traders have no longer been able to view, not even with delay, the operator that is actually trading or the category he belongs to. Then, this classification represents an ex-post sorting, as traders' identity is concealed to all the market participants.

To minimize the presence of outliers in the data, we exclude from the sample all the records where D_k is larger than 35 in absolute value. Moreover, we also eliminate the observations where at least one trader is the liquidity provider. Indeed, the liquidity provider may be called to trade because of contractual duties and his attribution to the category of informed or uninformed agents would be questionable. However, this choice should not affect the results, as it involves a very small amount of observations, sometimes even none[3].

3.4 The Empirical Specification

This section details the empirical specification and describes the methodology used to estimate the trader effect. We employ, as far as possible, a homogeneous specification for the whole sample, though each stock exhibits some specific peculiarities with respect to either the distribution of D_k or the significant lags of the explanatory variables.

3.4.1 Preliminary issues

As emphasized by HLM (1992), the first point to be examined concerns the number of m intervals used to classify the price changes. The objective is to find the optimal compromise between price resolution and estimation issues. Indeed, a large m turns out to be problematic for threshold estimation in extreme classes that collect only a

[3]For each stock, the maximum percentage of dropped observations is less than 0.05%. On the other hand, and to maintain uniformity within the data set, we fully exclude Dexia because of a significant number of transactions classified with '6'.

few observations. Conversely, selecting a small m makes price resolution unintelligible. Taking this trade-off into account, a feasible strategy to single out the optimal m consists of being driven by graphical and descriptive analysis. The aggregation of data in classes of D_k is performed according to the frequency plot and the distribution percentiles, with intervals that replicate the data dispersion around the central values, see Figure 3.1. For the whole sample, the number of m intervals is set equal to 5, 7 or 9, according to the distribution of D_k; a distinct class is always reserved to $D_k = 0$, as it includes most of the observations[4].

A second point to be discussed is related to the frequency used for sampling the data. The possible options are the clock-time convention, where data are selected according to a fixed sampling frequency (such as 5-minute intervals) or the event-time convention where all the transactions are included in the sample. In the last case, the inference is complicated by the fact that transaction times cannot be considered as independent and identically distributed. However, Easley et al. (1997) point out that clock-time stationarity in studies that examine information-based issues could seriously affect the results: a fixed sampling approach implies the loss of the information content included in the time pattern between market transactions. Since the main objective of this paper is to examine the impact on transaction prices caused by informed and uninformed traders, the transaction-time approach seems to be the natural choice. This is coherent with Hasbrouck (2007) who states that neither of the two approaches is preferable, and he suggests to set the sampling frequency according to the aim of the research.

3.4.2 Model regressors

For the empirical specification, the most relevant issue is regarding the explanatory variables used to express the impact of traders on transactions prices. For this purpose, we introduce a model regressor that combines the transaction volume and the trader cate-

[4]We also replicate the estimation by using different partitions for each stocks, and the results are only qualitatively affected.

Table 3.2: Transaction volumes and dummies for traders' identities. Data are from BNP.

	Buyer		Seller	
Volume	U	I	U	I
38	1	0	0	1
62	1	0	1	0
113	0	1	1	0
2,950	0	1	0	1

Table 3.3: Transaction volumes and trader combinations. Data are from BNP.

	Trader combinations			
Volume	BUSUVol	BUSIVol	BISUVol	BISIVol
38	0.000	38	0.000	0.000
62	62	0.000	0.000	0.000
113	0.000	0.000	113	0.000
2,950	0.000	0.000	0.000	2,950

gories. More precisely, for each observation we create four new variables according to the following steps:

- We generate two dummy variables, 'I' and 'U', that are equal to one if the transaction was executed by an informed-institutional trader or an uniformed-retail agent, respectively. These variables are defined for both sides of the market in order to specify whether the agent acts as a buyer or as a seller; obviously, the trader category of the two market sides may be the same or different.

- For each transaction, we multiply the volume by the trader dummies for the two market sides. In this way, we obtain four new variables which simultaneously express the traded volume and the types of investors behind each transaction.

As an example, consider Table 3.2 where we list four representative transactions from BNP; the table shows the volume, expressed as the number of exchanged shares, and the two dummy variables which identify the type of trader[5]. The first row of the table

[5]For the empirical application, we mean-normalize the transaction volume in order to moderate the scale effect on the coefficients.

displays a transaction volume equal to 38, the buyer is uninformed (U=1), and the seller is informed (I=1). The second row shows an exchanged volume of 62, with uninformed agents (U=1) on both market sides, and so on. By matching all the potential combinations of traders, we obtain Table 3.3, where BUSUVol refers to transactions in which both the buyer and the seller are uninformed, BUSIVol is for transactions where the buyer is uninformed and the seller is informed, BISUVol is for transactions where the buyer is informed and the seller uninformed, and BISIVol for transactions where both agents are informed traders. The first row of Table 3.3 displays the case of an uninformed buyer and an informed seller: BUSIVol is set equal to the traded volume, while the other variables are equal to zero. An analogous association is easily extended to the other three rows. It is immediately noticeable that only one combination of traders is responsible for the execution of each transaction, while the other combinations are marked as inactive, with zero volume. In this way, we emphasize the information content of volume, meanwhile associating a sort of sign to each observation. Under a broader perspective, this strategy reminds of the empirical papers where observed volume is connoted by the market direction (e.g. Hasbrouck , 1991). We label the impact of traders on transaction prices as the '*trader effect*', and we distinguish the four variables just described between *'cross trading'* and *'parallel trading'* variables. Cross trading is the circumstance where the operators on the two sides of the market are different (BUSIVol and BISUVol), while parallel trading (BUSUVol and BISIVol) considers the case of the same trader category.

The following regressors are included as controls in the mean specification:

- *Inter-trade durations* ($\Delta t_k = t_k - t_{k-1}$) expressed in seconds. The aim of this variable is to account for clock-time effects on the conditional mean of D_k^*. This is in line with HLM (1992), who use transaction-time events, but they allow for clock-time effects by including the inter-trade durations.

- *Seasonality.* The presence of an intraday pattern is modelled by means of a Fourier series according to the following sum

$$\sum_{i=1}^{p} cos(2\pi i\delta_k) + sin(2\pi i\delta_k).$$

where δ_k expresses the daily ratio between the time elapsed from 9.00 A.M. and the total duration of the continuous auction. The intraday seasonality has been especially noticed for volume or volatility patterns and it has been widely analyzed in the previous research, see Easley and O'Hara (1997) among others. Nevertheless, we also add a Fourier term in the conditional mean specification to account for possible nonlinearities in the evolution of price variations.

- *Init*, which identifies the sign of each transaction. In the empirical microstructure literature, several measures exist to determine the direction of a trade, i.e. to define if a transaction was initiated by a buyer or a seller. In this paper, we adopt the 'tick-test' algorithm proposed by Lee and Ready (1991), which represents the reference approach in this field. *Init* is equal to +1 when the transaction is buyer-initiated, and equal to -1 when it is seller-initiated.

3.4.3 Model specification and identification issues

This part discusses the exact specification of the mean and of the variance of the model, as well as the constraints required for a full identification of the parameter vector γ'. The optimal specification is selected according to model parsimony, parameters significance, and information criteria. We choose four lags of D_k, two lags of *Init*, a $p = 2$ for the seasonal component, and two lags of the four variables created to capture the trader effect. We adopt information criteria also to define the best specification for the trader effect in terms of lags to be accounted for, with possible alternatives between lags 1-2 or lags 2-3. We exclude contemporaneous values because of the endogeneity between volume and price. We limit to these two options, as including more than three lags is difficult to justify even in a high-frequency context, where algorithmic trading assures real time reactions to new information. Moreover, lags greater than three are often not significant anyway. The choice between the two couples of lags is mainly dictated by the better fit of the model indicated by the information criteria; however, the estimates are generally similar even under the discarded option.

Finally, the issue relative to identification constraints is strictly related to the variance specification. In the basic model, we employ a variance normalization that assumes no explicit design for σ_k. Without any restriction, it is impossible to achieve the model identification, provided that there exist multiple combinations of the parameters which leave the likelihood unchanged. When an explicit form of heteroskedasticity is not taken into account, model identification is achieved by excluding the constant from the list of the regressors, or by fixing a threshold, α_j. We prefer to exclude the constant, in line with the discussion on interval regression. The identification constraints get slightly more complicated whenever one decides to consider a set of explanatory variables affecting the variance σ_k. Indeed, the inclusion of the scale dimension increases the number of parameter combinations that could generate the same value of the objective function. In this case, the issue of identification is solved by fixing two thresholds or by excluding the constant both from the mean and from the variance of the error term. This last option is the one adopted in Section 3.7, where some alternative specifications are considered as a robustness test.

3.5 Results

This section analyses the ML estimates of the basic ordered probit model, and all the following discussion is based on a significance level of 1%. The choice of such a threshold is motivated by the size of the data set used in this research and it is not a minor point. In fact, it becomes easier to reject any null hypothesis when the number of observations is so impressive; however, our findings are generally confirmed even at lower significance thresholds. Table 3.4 displays the estimates for two representative stocks, Bouygues and

Table 3.4: ML estimates of the ordered probit model, with p-values in parentheses. The table shows the results for Bouygues and BNP. The lags used for the estimation of the trader effect are indicated in brackets.

	Bouygues		BNP	
Variable	Estimate[1,2]	P-value	Estimate[2,3]	P-value
D_{k-1}	-4.01e-02	(0.00)	-4.41e-02	(0.00)
D_{k-1}	-3.48e-02	(0.00)	-3.48e-02	(0.00)
D_{k-3}	-1.51e-02	(0.00)	-1.24e-02	(0.00)
D_{k-4}	-4.43e-03	(0.00)	-5.11e-03	(0.00)
Δt_k	-7.89e-04	(0.00)	5.78e-04	(0.03)
$Init_{k-1}$	-6.70e-02	(0.00)	-6.90e-02	(0.00)
$Init_{k-2}$	6.79e-02	(0.00)	6.65e-02	(0.00)
$Cos(2\pi\delta)$	-5.44e-03	(0.08)	1.93e-03	(0.31)
$Cos(4\pi\delta)$	-4.23e-03	(0.16)	-1.92e-03	(0.31)
$Sin(2\pi\delta)$	-4.36e-03	(0.15)	6.06e-04	(0.75)
$Sin(4\pi\delta)$	-7.29e-03	(0.00)	-1.91e-03	(0.31)
BUSUVol(k-i)	3.16e-03	(0.11)	4.19e-04	(0.72)
BUSUVol(k-j)	-1.94e-03	(0.35)	-5.25e-04	(0.66)
BUSIVol(k-i)	-1.30e-02	(0.00)	-7.29e-03	(0.00)
BUSIVol(k-j)	-1.64e-02	(0.00)	-8.75e-03	(0.00)
BISUVol(k-i)	2.24e-02	(0.00)	1.20e-02	(0.00)
BISUVol(k-j)	1.72e-02	(0.00)	1.26e-02	(0.00)
BISIVol(k-i)	1.54e-03	(0.44)	1.00e-04	(0.93)
BISIVol(k-j)	-1.71e-03	(0.39)	-6.95e-04	(0.57)

BNP. We select these two stocks to illustrate the two lag alternatives, but the results can be generalized to the rest of the sample. The complete set of estimates is available upon request. The main findings are summarized as follows:

- At least three of the four lags of D_k included in the mean are negative and statistically significant. This appears to be a general outcome which holds for the whole sample. This result is not unexpected and it reflects the pattern displayed by the correlogram of D_k; as highlighted in previous studies (e.g. Roll, 1984), this is consistent with the occurrence of reversals in transaction prices.

- The interpretation of Δt_k is not immediate as there is no homogeneous outcome in the whole sample. Generally speaking, it is hard to clearly outline the direction of duration effects on the conditional mean. This is evident in Table 3.4, where

Bouygues exhibits a negative coefficient for Δt_k, while for BNP it becomes positive, though not statistically significant at 1%. These results can be extended to the full data set, and they match the ones provided in HLM (1992) who also find an ambiguous impact for clock-time effects.

- Table 3.4 shows that the coefficients of the Fourier series are almost never significant at 1%. A LR test to fully exclude the seasonal component from the conditional mean rejects the null hypothesis only for a limited number of cases (e.g. Bouygues, Vinci, or Crédit Agricole). On the whole, the seasonality effect for price variations seems definitely weak, at least with respect to the intraday pattern which is traditionally observed for trade durations or transaction volume.

- The variable *Init* always displays a negative coefficient for the first lag, and either a positive or a non-significant coefficient for the second lag. The inclusion of *Init* in the list of regressors should account for the effect of the bid-ask bounce, and in HLM (1992) all the lags of this variable exhibit a negative coefficient. This discrepancy could be attributed to the different approaches adopted to classify the transactions. HLM (1992) employ an indicator variable that discriminates using quote data, while we adopt the tick-test algorithm, since only the transaction data are available in our sample. Actually, this could reduce the capacity of *Init* to capture the bid-ask bounce.

- As to trader effect variables, Table 3.4 presents a homogeneous outcome for the two stocks, which can be generalized to the whole sample. Table 3.4 shows that cross trading is always significant at 1%, while parallel trading exhibits no significant estimates for both lags. However, it is particularly interesting to note the sign displayed by BUSIVol and BISUVol. According to our estimates, when an informed agent sells to an uninformed one, the effect on the conditional mean is negative; the opposite holds when an informed trader buys from an uninformed one. Table 3.5 summarizes these findings for the whole sample: the rows indicate the number

Table 3.5: Whole sample results for trader effect variables. The table provides the percentage of stocks falling in each row-column combination.

# Signif. Lags	BUSUVol	BUSIVol	BISUVol	BISIVol
0	97.44	2.56	7.69	89.74
1	2.56	10.26	7.69	10.26
2	0	87.18	84.62	0

of significant lags, while the columns report the trader effect variables. The cells report the percentage of stocks falling in each row-column combination; obviously, the table is built by considering the case of a negative coefficient for BUSIVol and a positive one for BISUVol. For parallel trading, Table 3.5 shows that both lags of BUSUVol are never significant in 97.44% of the sample, while BISIVol is not significant for both lags in 89.74% of the cases, and it exhibits one significant coefficient only for a small fraction of the series. On the other hand, cross trading displays the opposite result. Indeed, more than 80% of the stocks presents a significant coefficient for both lags of BISUVol, and this percentage almost reaches the 90% in the case of BUSIVol; additionally, the fraction of stocks in which both lags of cross trading are not significant is generally marginal[6]. Altogether, the estimates in Tables 3.4 and 3.5 provide evidence of a trader-related effect, with stock prices following the direction of institutional trading. By reasonably assuming that institutional investors benefit from a large amount of information (e.g. Chakravarty, 2001), our results are coherent with a price effect generated by informed-based trading. Quite interestingly, we find no significant result for BISIVol; in this case, it seems that the market does not single out a prevailing impact between the two institutional pressures. On this basis, it is not surprising to get non-significant estimates when both traders are uninformed, as these investors are mostly liquidity-motivated. In terms of market efficiency, these findings show that private information is rapidly incorporated into market prices. The effect does not only involve the conditional

[6]EdF is the only stock with two non-significant lags of BUSIVol (2.5%), while Alcatel, Ppr, and STM are the three stocks with two non-significant lags of BISUVol (7.69%).

mean, but the full conditional distribution, as will be discussed in Section 3.6. Generally speaking, these results may appear puzzling since the traders' identity is concealed to all market operators. Even though the market 'absorbs' the behaviour of informed traders, the resulting impact cannot be explicitly driven by an imitating strategy. Then, different sources of information should be useful to detect the trading by institutional agents (e.g. clustering of transactions, transaction volume), and we postpone this issue to Section 3.8, where it is examined further.

3.6 Postestimation results

This section concentrates on postestimation issues for the basic ordered probit. Particular attention is given to test the presence of serial dependence in the residuals, in order to evaluate the dynamic specification of the model. Moreover, we also discuss the marginal effects associated with the traders' activity. In particular, we examine the marginal response probabilities for the ordered probit model, and we simulate the price impact under different market scenarios.

3.6.1 Autocorrelation issues

The investigation of residual diagnostics is not immediate in the case of latent variable models, as the dependent variable is not observed. The approach that is usually adopted refers to the concept of generalized residuals defined in Gourieroux et al. (1985). In the case of the ordered probit model, given that $D_k = d_j$, the generalized residuals $\hat{\epsilon}_k$ are computed as:

$$\begin{aligned}
\hat{\epsilon}_k &= E[\epsilon_k | D_k = d_j, X_k, W_k; \hat{\gamma}] = \hat{\sigma}_k \frac{\phi(c_1) - \phi(c_2)}{\Phi(c_2) - \Phi(c_1)}, \\
c_1 &= \frac{1}{\hat{\sigma}_k} \left(\hat{\alpha}_{j-1} - \mathbf{X}_\mathbf{k}' \hat{\beta} \right) \\
c_2 &= \frac{1}{\hat{\sigma}_k} \left(\hat{\alpha}_j - \mathbf{X}_\mathbf{k}' \hat{\beta} \right).
\end{aligned}$$

where $\hat{\gamma}$ is the ML estimation of the parameters, and ϕ represents the standard normal probability density function. From the previous formula, it is straightforward to compute a test that verifies the presence of autocorrelation in the residuals, and a full description about the score statistics of interest can be found in HLM (1992). Under the null hypothesis of no serial correlation, the score statistics has a χ^2_1 distribution:

$$\hat{\xi}_j = \frac{\left(\sum_{k=j+1}^{n} \hat{D}_{k-j}\hat{\epsilon}_k\right)^2}{\sum_{k=j+1}^{n} \hat{D}^2_{k-j}\hat{\epsilon}^2_k} \tag{3.3}$$

$$\hat{D}_k = X'_k\hat{\beta} + \hat{\epsilon}_k.$$

Equation 3.3 can be used to test any order j of serial correlation in the residuals, and it keeps the same number of degrees of freedom, regardless of the value of j. Table 3.6 displays the values of $\hat{\xi}_j, j = 1, ..., 8$ for Bouygues and BNP. For $j = 1, ..., 4$, the table shows that $\hat{\xi}_j$ is always less than 6.635, which represents the critical value at 1% for a χ^2_1 distribution. After the fourth lag, the outcome of the score statistics is not as uniform, though we generally find a rejection of the null hypothesis of no serial correlation. This result is not surprising if associated with the presence of four lags of D_k in the mean specification, as HLM (1992) similarly pointed out. In considering the whole sample, a joint test for the absence of autocorrelation at lag 4 is rejected for almost half of the stocks. HLM (1992) obtain similar findings, even though our research uses a number of observations that is considerably larger. The size of the data set could represent a first explanation for the failure to reject the null hypothesis, suggesting that a smaller significance level is more appropriate. Alternatively, this finding could be attributed to the excessively limited dynamics of the estimated model, and Section 3.7 checks for this possibility. We compute the score statistics in the case of an *extended* probit model, considering a larger number of lags for all the mean regressors, as well as powers of the trader effect variables. However, even in this case, almost half of the sample still rejects the null of no serial correlation for the first 4 lags. Moreover, the small benefit observed

Table 3.6: Score statistics $\hat{\xi}_j$ for the null hypothesis of no serial correlation in the ordered probit disturbances.

Lags	Bouygues	BNP
$\hat{\xi}_1$	0.245	0.925
$\hat{\xi}_2$	2.300	5.081
$\hat{\xi}_3$	6.105	3.687
$\hat{\xi}_4$	4.570	1.456
$\hat{\xi}_5$	6.405	19.842
$\hat{\xi}_6$	8.388	25.431
$\hat{\xi}_7$	17.533	24.996
$\hat{\xi}_8$	2.173	44.843

in terms of less autocorrelated residuals is generally overwhelmed by the loss of model parsimony, and the extended probit is usually rejected by a LR test with respect to the basic specification.

3.6.2 Marginal effects

For the basic ordered probit, we present the change in the response probability as marginal effects. Indeed, when the dependent variable has not a quantitative value, the marginal response probability is more appropriate than the marginal effect on the conditional mean. This is the case of a discrete dependent variable model as the ordered probit, for which we compute the marginal response probabilities as:

$$\frac{\partial p(D = d_j|\bar{\mathbf{X}}, \hat{\beta})}{\partial X_v} = \begin{cases} -\hat{\beta}_v \phi(\alpha_j - \bar{\mathbf{X}}'\hat{\beta}) & \text{if } j = 1, \\ \hat{\beta}_v \phi(\alpha_j - \bar{\mathbf{X}}'\hat{\beta}) & \text{if } j = m, \\ \hat{\beta}_v [\phi(\alpha_{j-1} - \bar{\mathbf{X}}'\hat{\beta}) - \phi(\alpha_j - \bar{\mathbf{X}}'\hat{\beta})] & \text{if } 1 < j < m \end{cases} \tag{3.4}$$

where $\bar{\mathbf{X}}$ represents the vector of means of the model regressors, X_v stands for a generic $v - th$ explanatory variable, and $\hat{\beta}_v$ serves as the corresponding estimated coefficient.

For a marginal variation in any of the regressors, the marginal response probabilities

Table 3.7: Marginal response probabilities for Bouygues. The table displays the estimates for the trader effect variables; p-values are reported in parentheses. The headers indicate the $j - th$ interval of the D_k distribution.

	1st Interv.		2nd Interv.		3rd Interv.		4th Interv.	
	Marg.	P-value	Marg.	P-value	Marg.	P-value	Marg.	P-value
BUSUVol(t-1)	-2.51E-04	(0.11)	-2.93E-04	(0.11)	-3.00E-04	(0.11)	1.28E-05	(0.14)
BUSUVol(t-2)	1.54E-04	(0.35)	1.80E-04	(0.35)	1.84E-04	(0.35)	-7.85E-06	(0.36)
BUSIVol(t-1)	1.04E-03	(0.00)	1.21E-03	(0.00)	1.24E-03	(0.00)	-5.29E-05	(0.00)
BUSIVol(t-2)	1.31E-03	(0.00)	1.53E-03	(0.00)	1.56E-03	(0.00)	-6.67E-05	(0.00)
BISUVol(t-1)	-1.78E-03	(0.00)	-2.08E-03	(0.00)	-2.13E-03	(0.00)	9.09E-05	(0.00)
BISUVol(t-2)	-1.36E-03	(0.00)	-1.59E-03	(0.00)	-1.63E-03	(0.00)	6.96E-05	(0.00)
BISIVol(t-1)	-1.22E-04	(0.44)	-1.43E-04	(0.44)	-1.46E-04	(0.44)	6.23E-06	(0.45)
BISIVol(t-2)	1.36E-04	(0.39)	1.59E-04	(0.39)	1.63E-04	(0.39)	-6.93E-06	(0.40)

	5th Interv.		6th Interv.		7th Interv.	
	Marg.	P-value	Marg.	P-value	Marg.	P-value
BUSUVol(t-1)	2.87E-04	(0.11)	2.91E-04	(0.11)	2.53E-04	(0.11)
BUSUVol(t-2)	-1.76E-04	(0.35)	-1.78E-04	(0.35)	-1.55E-04	(0.35)
BUSIVol(t-1)	-1.19E-03	(0.00)	-1.20E-03	(0.00)	-1.05E-03	(0.00)
BUSIVol(t-2)	-1.50E-03	(0.00)	-1.51E-03	(0.00)	-1.32E-03	(0.00)
BISUVol(t-1)	2.04E-03	(0.00)	2.06E-03	(0.00)	1.80E-03	(0.00)
BISUVol(t-2)	1.56E-03	(0.00)	1.58E-03	(0.00)	1.38E-03	(0.00)
BISIVol(t-1)	1.40E-04	(0.44)	1.42E-04	(0.44)	1.23E-04	(0.44)
BISIVol(t-2)	-1.56E-04	(0.39)	-1.57E-04	(0.39)	-1.37E-04	(0.39)

measure the change in the probability of observing a specific outcome d_j. Table 3.7 shows the marginal response probabilities for Bouygues, but the results can be extended in a similar way to the whole sample and they are available upon request. We only consider the variables of interest for this research, i.e. the ones which refer to the trader effect. Table 3.7 consists of seven columns, one for each interval used to classify the price variations of Bouygues. Clearly, the number of marginal response probabilities to be computed depends on the intervals used to partition the frequency distribution of D_k. In the case of Table 3.7, the first three columns consider the marginal effect for negative price variations, column 4 refers to null price variations, while the last three columns are devoted to positive values of D_k. According to Table 3.7, the response probability is not significant for parallel trading across all the seven intervals, which confirms the estimation results provided in Table 3.4. When parallel trading occurs, there is no significant impact on the direction of the trading process. This supports the conclusion that the market does not recognize an informative signal when both traders are uninformed, or it is unable to distinguish a leading trading path when both buyer and seller are informed agents. Conversely, the marginal effect for cross trading is significant, but with opposite patterns for BUSIVol and BISUVol. BUSIVol exhibits a positive sign in the first classes which collect negative price variations, while the marginal effect becomes negative for intervals that include positive price variations; the opposite pattern is observed in the case of BISUVol. The implication in terms of market direction is coherent with the previous analysis. When an informed trader buys, the probability of a positive price variation augments in the following transactions; meanwhile, the probability of a negative price change decreases. Clearly, the reverse conclusion is valid when informed traders sell to an uninformed buyer. As to the central class, the marginal effect for cross trading is not uniform in the whole sample; however, it is generally small or even not significant. On the whole, the estimates confirm the presence of a significant trader effect, with the market moving along the trading direction of informed agents.

The previous analysis is more easily appreciated by looking at Figures 3.3 and 3.4.

Figure 3.3: Plot of marginal response probabilities for lagged values of BUSUVol and BISIVol, across distribution intervals of D_k. The plots are referred to Bouygues. The central solid line represents the estimated marginal effect, while the two dashed lines define confidence intervals at 99%.

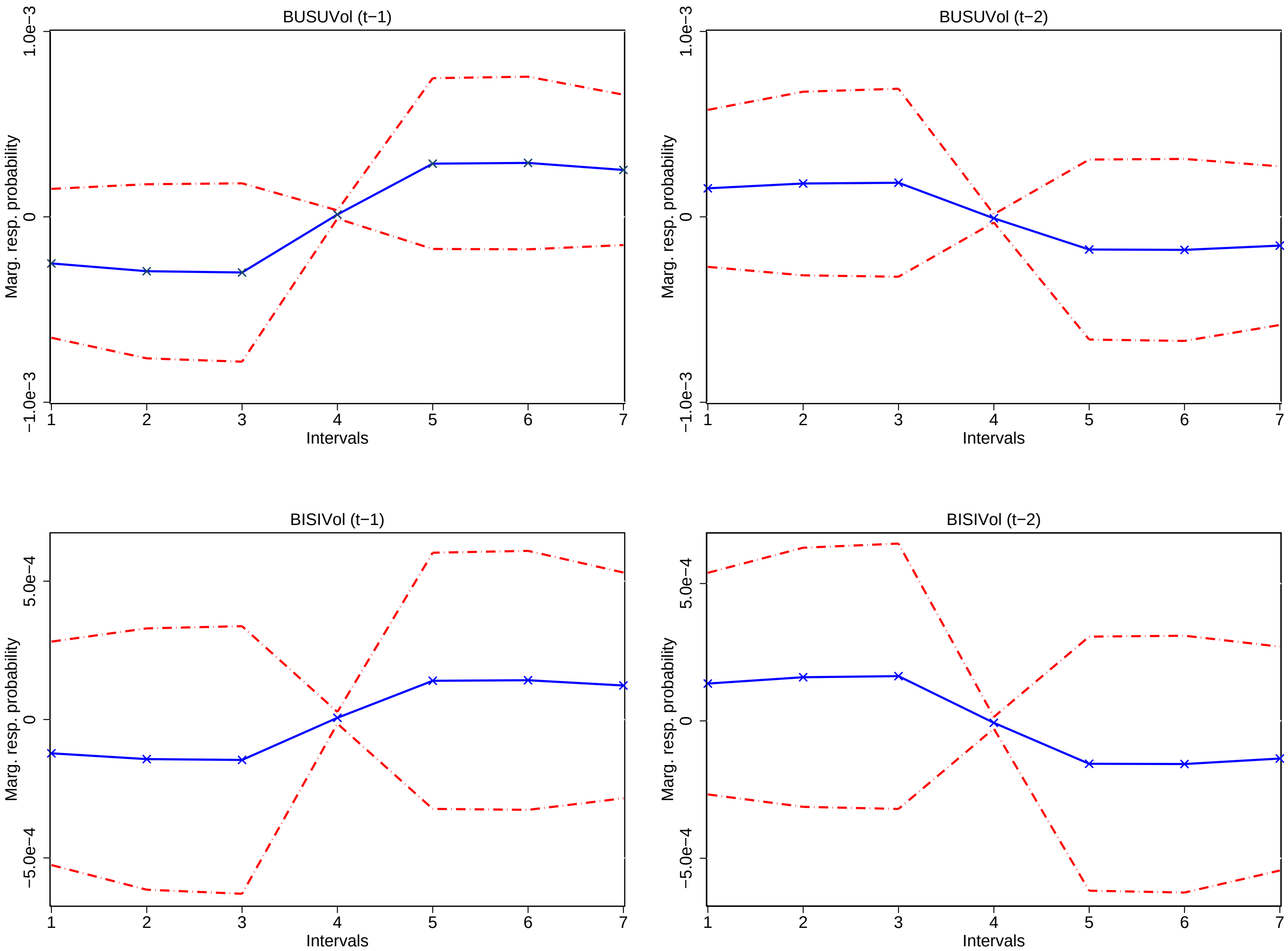
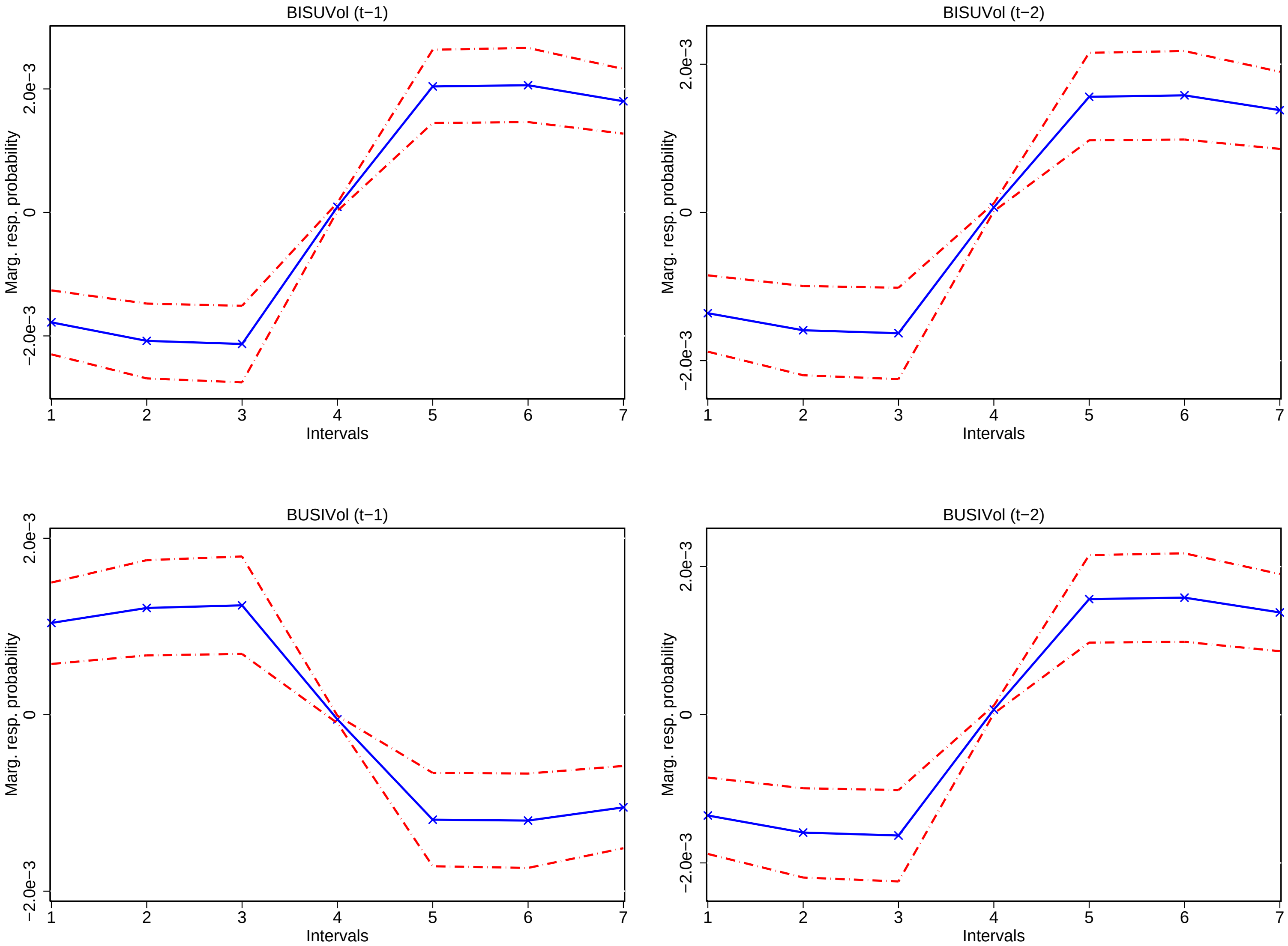


Figure 3.4: Marginal response probabilities for lagged values of BISUVol and BUSIVol, across distribution intervals of D_k. The plots are referred to Bouygues. The central solid line represents the estimated marginal effect, while the two dashed lines define confidence intervals at 99%.



The former plots the marginal response probabilities for BUSUVol and BISIVol; the latter displays the same statistics for BUSIVol and BISUVol. The four panels in Figure 3.3 shows that the marginal effect is never significant, and it fluctuates between wide confidence intervals. On the other hand, in the case of cross trading, the marginal effects are estimated quite accurately and they exhibit an opposite swinging path in correspondence of $D_k = 0$. From Figure 3.4, we also notice that the response probability computed for the class $D_k = 0$ is very close to zero, but still significant, at least for Bouygues. Although the marginal effects could appear fairly small in terms of the size of the impact, we recall that our estimates are based on mean-normalized transaction volumes, and the marginal effects represent probability variations, which partially explains the small impact. Moreover, the response probabilities are obtained by considering just a few lags of marginal variation in the explanatory variables. Looking at the average number of daily transactions in Table 3.1, it would be unreasonable to expect a sizeable impact, even in the case of large transactions.

3.6.3 Price impact analysis

Besides the response probabilities, we also examine the price impact as the effect of a current transaction on the conditional distribution of the following price change. We adopt the approach described in HLM (1992), i.e. we assume some specific values of the model regressors and we compute the conditional probabilities of observing $D_k = d_j$. Then, we use the conditional probabilities to evaluate the impact of a particular trading sequence both on the conditional mean and the distribution of D_k. To proceed, we set Δt_k and the variables of the Fourier series to their sample mean, while for lag of D_k and for *Init* we consider three alternative scenarios, as defined in Table 3.8. We follow a similar strategy for BUSUVol, BUSIVol, BISUVol and BISIVol, bearing in mind that only one combination is different from zero for each transaction. Without loss of generality, we consider a sequence of two lags of the same regressor, and we assume five distinct values from the distribution of transaction volume, corresponding to the 5th, 25th, 50th, 75th,

Table 3.8: Scenarios for the price impact analysis.

Variable	Scenario 1	Scenario 2	Scenario 3
D_{k-1}	0	2	1
D_{k-2}	-1	0	-1
D_{k-3}	0	0	0
D_{k-4}	1	-3	0
$Init_{k-1}$	-1	1	-1
$Init_{k-2}$	1	-1	-1

Table 3.9: Distribution percentiles of transaction volume. The table reports the number of shares.

	Percentile				
Stock	5th	25th	50th	75th	95th
Bouygues	15	84	190	302	804
BNP	16	97	200	400	1,083

and 95th percentile; Table 3.9 reports the percentiles for Bouygues and BNP. Clearly, the data set does not generally display two consecutive lags of the same regressor with an identical transaction volume; however, this is just a simplifying assumption for illustrative purposes. Indeed, we also try different values and alternative combinations of the lags of BUSUVol, BUSIVol, BISUVol, and BISIVol. The price impact of cross trading is still evident, though it is obviously less remarkable for a small transaction size or for a combination of lags between cross trading and parallel trading. Table 3.10 presents the effect on the conditional mean evaluated under the three scenarios. Each column provides the price impact as the conditional mean difference with respect to the 5th percentile; the difference is expressed as the percentage of the average transaction price displayed in Table 3.1. As an example, the first entry of Table 3.10 is 0.001 which corresponds to $\frac{E[D_k^{25th}]-E[D_k^{5th}]}{45.72} \cdot 100$. Table 3.10 confirms the results discussed in Section 3.5 for cross and parallel trading. Informed traders affect the conditional mean when the counterpart of the exchange is an uninformed agent. The sign of the impact follows the direction of institutional trading, while the size of the effect increases along with transaction volume. On the other hand, the effect of parallel trading is very small, even for a large volume

Table 3.10: Price impact on the conditional mean. The table shows the conditional mean difference with respect to the 5th percentile; the difference is expressed as the percentage of the average transaction price. The column header indicates the trader effect variable used for the price impact analysis.

	Scenario 1							
	Bouygues				BNP			
	BUSUVol	BUSIVol	BISUVol	BISIVol	BUSUVol	BUSIVol	BISUVol	BISIVol
25th perc.	0.001	-0.017	0.023	0.000	0.000	-0.006	0.010	0.000
50th perc.	0.002	-0.044	0.059	0.000	0.000	-0.014	0.022	-0.001
75th perc.	0.003	-0.071	0.096	0.000	0.000	-0.030	0.046	-0.001
95th perc.	0.008	-0.196	0.266	-0.001	-0.001	-0.083	0.128	-0.003

	Scenario 2							
	Bouygues				BNP			
	BUSUVol	BUSIVol	BISUVol	BISIVol	BUSUVol	BUSIVol	BISUVol	BISIVol
25th perc.	0.001	-0.017	0.023	0.000	0.000	-0.006	0.010	0.000
50th perc.	0.002	-0.044	0.059	0.000	0.000	-0.014	0.022	-0.001
75th perc.	0.003	-0.072	0.096	0.000	0.000	-0.030	0.046	-0.001
95th perc.	0.008	-0.198	0.264	-0.001	-0.001	-0.084	0.128	-0.003

	Scenario 3							
	Bouygues				BNP			
	BUSUVol	BUSIVol	BISUVol	BISIVol	BUSUVol	BUSIVol	BISUVol	BISIVol
25th perc.	0.001	-0.017	0.023	0.000	0.000	-0.006	0.010	0.000
50th perc.	0.002	-0.043	0.058	0.000	0.000	-0.014	0.022	-0.001
75th perc.	0.003	-0.071	0.096	0.000	0.000	-0.030	0.046	-0.001
95th perc.	0.008	-0.195	0.263	-0.001	-0.001	-0.083	0.127	-0.003

size. Actually, we recognize that the price impact is generally small for cross trading too. Nevertheless, the magnitude of our estimates seems in line with the results reported in HLM (1992); moreover, the previous remark on the size of marginal response probabilities also applies in this case. Interestingly, the results reported in Table 3.10 appear mostly influenced by the trader effect variables. In fact, under the same combination of traders, the estimates are quite similar across the three scenarios.

Finally, we display in Figure 3.5 and 3.6 the price impact on the whole conditional distribution. The two figures provide the estimated conditional probabilities for Bouygues and BNP, and we reduce the scale of the y-axis to make the effect on the tails of the distribution more appreciable. Actually, this is not a big loss as the price impact for the central interval is almost negligible, in line with the previous discussion for marginal effects. To emphasize our results, we show the conditional probabilities computed for a transaction volume equal to the 5th and the 95th percentile. We focus on these percentiles as they represent two opposite cases in terms of transaction size. Nevertheless, the principle of our analysis also extends to other trading volumes, though the magnitude of the effect is clearly smaller. The upper plots of Figure 3.5 and 3.6 display the price impact for BUSUVol and BISIVol; the effect is proven to be extremely small, independently of the transaction size. Conversely, when cross trading is taken into account, we observe a shift in the conditional distribution of price variations. The impact exhibits the usual asymmetric pattern for BUSIVol and BISUVol, and it is particularly evident in the tails of the distribution. When informed traders sell, an increasing volume moves the conditional distribution to the left, i.e. towards negative price variations. The opposite holds when an informed traders buy from an uniformed agent; in this case, a higher volume shifts the conditional distribution towards positive values of D_k.

Figure 3.5: Price impact of a simulated trading pattern (Scenario 1). The graphs report the estimated ordered probit conditional probabilities. The green bar is referred to the 5th percentile of the transaction volume distribution, while the orange bar is referred to the 95th percentile.

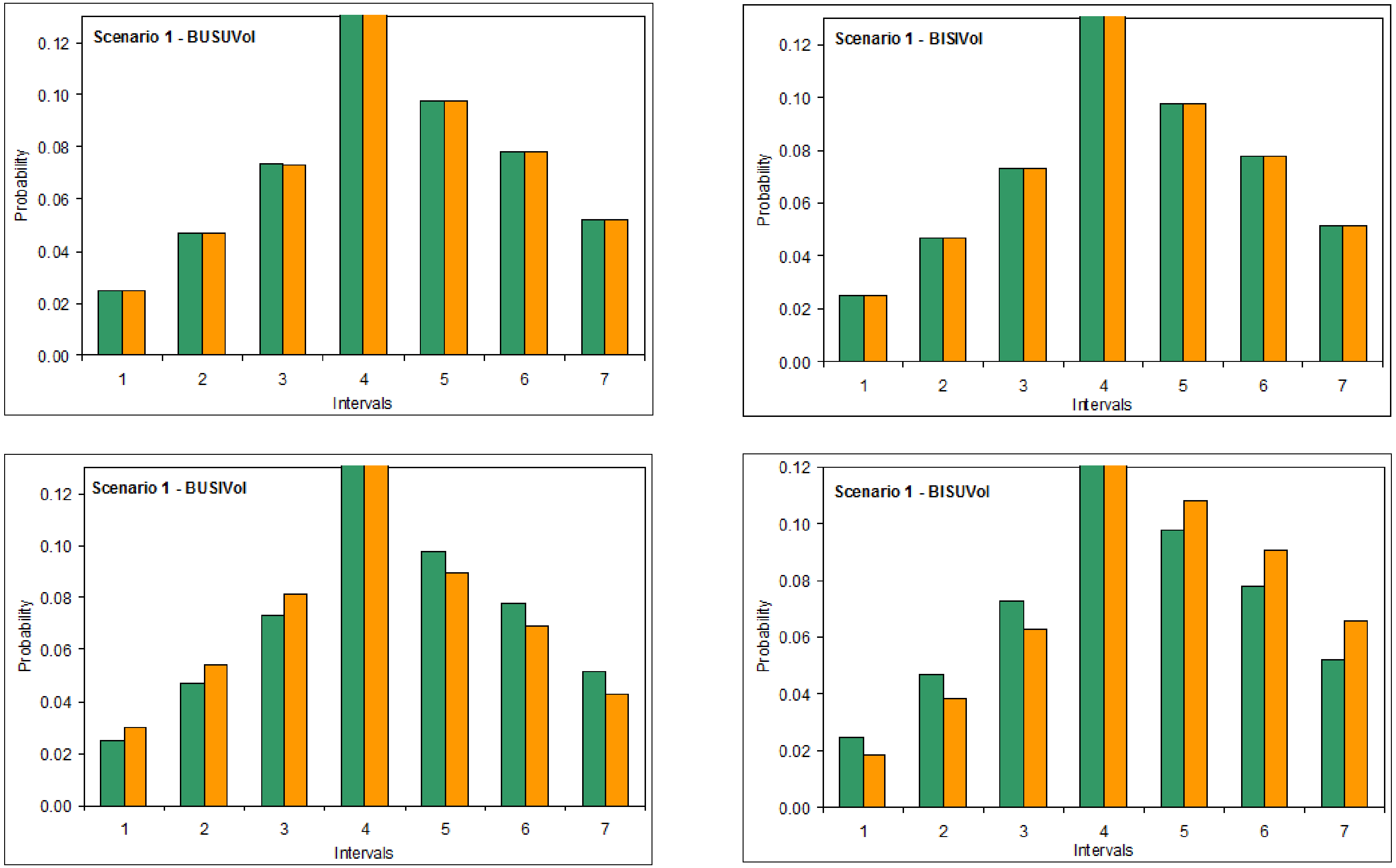
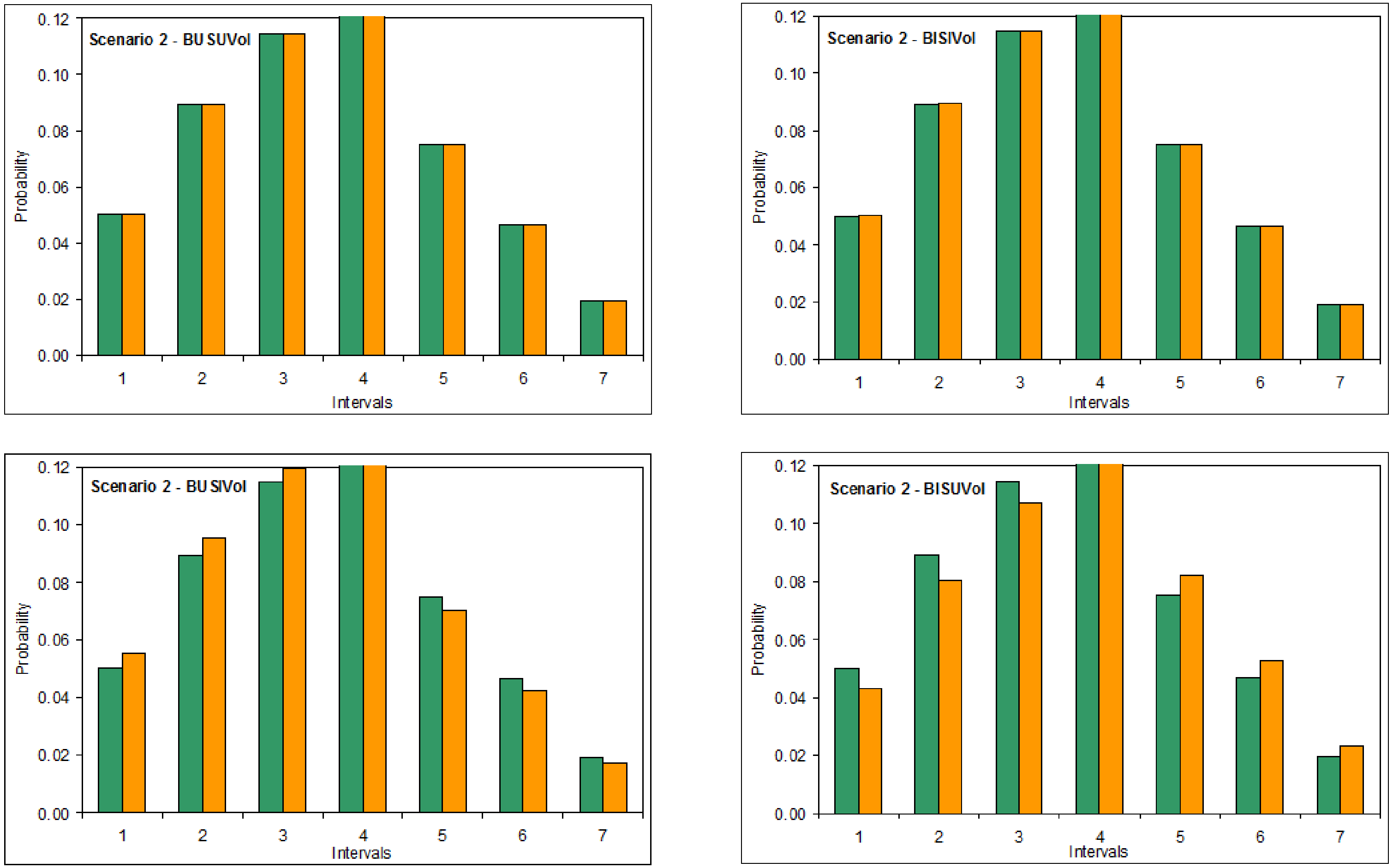


Figure 3.6: Price impact of a simulated trading pattern (Scenario 2). The graphs report the estimated ordered probit conditional probabilities. The green bar is referred to the 5th percentile of the transaction volume distribution, while the orange bar is referred to the 95th percentile.



3.7 Robustness Tests

In this section, we examine three alternative empirical specifications to test the robustness of our findings. More precisely, we re-estimate the model using an interval regression, an ordered probit with an *extended* set of regressors, and an ordered probit with an explicit form of heteroskedasticity. We take out OLS from the set of potential alternatives, the motivation for which is twofold (see HLM, 1992). First, using OLS is equivalent to assuming that the dependent variable has a continuous attribute, thus neglecting the presence of price discreteness. Second, unlike the ordered probit, OLS does not capture the nonlinearities present in the data. All the three specifications are only employed for robustness purposes, and they are disregarded as models of reference. Indeed, the LR test and the information criteria point to the basic ordered probit being preferable in terms of model parsimony. Furthermore, when an explicit form of heteroskedasticity is taken into account, the estimation time sensibly increases, though no quantitatively relevant impact is encountered with respect to the basic findings.

The interval regression model simply replicates the original ordered probit specification, but it excludes the partition thresholds from the vector of parameters to be estimated. With respect to the basic model, the *extended* ordered probit includes seven lags for D_k, two lags for Δt_k, an additional lag for the variables which measure the trader effect, together with the corresponding squared values. The resulting mean specification for a representative stock is[7]:

$$X_k\beta = \sum_{i=1}^{7} \beta_i D_{k-i} + \beta_8 \Delta t_k + \beta_9 \Delta t_{k-1} + \beta_{10} Init_{k-1} + \beta_{11} Init_{k-2} +$$
$$\beta_{12} cos(2\pi\delta_k) + \beta_{13} cos(4\pi\delta_k) + \beta_{14} sin(2\pi\delta_k) + \beta_{15} sin(4\pi\delta_k) +$$
$$\sum_{i=0}^{2} \beta_{16+i} BUSUVol_{k-i-1} + \sum_{i=0}^{2} \beta_{19+i} BUSIVol_{k-i-1} + \sum_{i=0}^{2} \beta_{22+i} BISUVol_{k-i-1} +$$

[7]The following formula refers to the case of lags [1,2]. It is immediate to extend it to the case of lags [2,3].

$$\sum_{i=0}^{2}\beta_{25+i}BISIVol_{k-i-1}+\sum_{i=0}^{2}\beta_{28+i}BUSUVol^2_{k-i-1}+\sum_{i=0}^{2}\beta_{31+i}BUSIVol^2_{k-i-1}+$$
$$\sum_{i=0}^{2}\beta_{34+i}BISUVol^2_{k-i-1}+\sum_{i=0}^{2}\beta_{37+i}BISIVol^2_{k-i-1.} \quad (3.5)$$

Equation (3.5) displays a considerable number of parameters to be estimated, which contrasts with the principle of model parsimony. However, the inclusion of additional regressors is motivated by robustness purposes. More precisely, we expect the trader effect to persist even when the mean specification is enriched with additional lags or with the second power of the explanatory variables.

The last specification employed as robustness check considers the following structure for the variance of the error distribution:

$$\begin{aligned} W_k\theta = \theta_1 D_{k-1} + \theta_2 D_{k-2} + \theta_3 \Delta t_k + \\ \theta_4 cos(2\pi\delta_k) + \theta_5 cos(4\pi\delta_k) + \theta_6 sin(2\pi\delta_k) + \theta_7 sin(4\pi\delta_k). \end{aligned} \quad (3.6)$$

Equation (3.6) serves to verify that the inclusion of a scale factor only affects the magnitude of the estimates, keeping the direction of the marginal effect unchanged[8]. We include a seasonal component defined as the Fourier series adopted for the mean; this helps to recover the stylized pattern displayed by the intraday volatility. The presence of Δt_k should account for clock-time effects in the variance, and similarly the lagged values of D_k control for the impact of price variations. We consider the possibility of testing the trader effect also in the variance, but some preliminary estimates show the absence of a clear and unambiguous outcome. Moreover, the estimates are often non-significant, suggesting that the asymmetric effect related to trader identity only influences the conditional mean.

Table 3.11 shows the estimates for the three specifications discussed in this section, limitedly to the trader effect variables. We still restrict the exposition to Bouygues and

[8]STATA OGLM routine developed by Richard Williams is employed to achieve ML estimates for this model specification. The corresponding estimates are labelled as OGLM.

Table 3.11: Estimates of the model specifications adopted for robustness check; p-values are in parentheses. "OGLM" refers to the heteroskedastic ordered probit, "Interval" to interval regression, and "Extended" to the extended ordered probit. The table only displays the trader effect variables (the first two lags for the extended ordered probit model). The lags used for the estimation are indicated in brackets.

	Bouygues[1,2]			BNP[2,3]		
	OGLM	Interval	Extended	OGLM	Interval	Extended
BUSUVol(t-i)	2.80e-03	3.82e-03	3.50e-03	4.26e-04	9.71e-04	-8.19e-04
	(0.21)	(0.19)	(0.25)	(0.74)	(0.55)	(0.59)
BUSUVol(t-j)	-1.48e-03	-1.93e-03	-3.74e-03	-1.07e-03	-1.39e-03	-6.07e-05
	(0.50)	(0.50)	(0.22)	(0.40)	(0.39)	(0.97)
BUSIVol(t-i)	-1.02e-02	-2.34e-02	-1.57e-02	-5.19e-03	-1.17e-02	-1.03e-02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BUSIVol(t-j)	-1.60e-02	-2.37e-02	-2.04e-02	-7.81e-03	-1.36e-02	-1.18e-02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BISUVol(t-i)	1.99e-02	3.66e-02	3.02e-02	9.82e-03	1.90e-02	1.55e-02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BISUVol(t-j)	1.57e-02	2.50e-02	2.04e-02	1.21e-02	1.97e-02	1.44e-02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BISIVol(t-i)	1.49e-03	2.62e-03	4.25e-03	3.33e-04	7.34e-05	2.02e-03
	(0.48)	(0.35)	(0.10)	(0.80)	(0.97)	(0.20)
BISIVol(t-j)	-1.91e-03	-2.61e-03	-2.71e-03	-4.87e-04	-9.64e-04	-1.13e-03
	(0.36)	(0.35)	(0.30)	(0.70)	(0.57)	(0.48)

BNP, though our findings extend to the whole sample. Table 3.11 displays the parameter estimates for OGLM and interval regression, while the marginal effect on the conditional mean is provided for the extended probit. We differentiate the extended probit model because it also includes squared variables, so displaying only the coefficients of first-order variables is essentially meaningless. Although the estimates are not comparable in terms of magnitude, Table 3.11 shows full uniformity with the results displayed in Table 3.4. For all the three specifications, the sole cross trading is significant, with a negative sign for BUSIVol and a positive sign for BISUVol. On the contrary, parallel trading is never significant, which reinforces the results discussed in Section 3.5 for the basic model. In the *extended* probit model, cross trading variables generally present a significant estimate also in the additional lag (not displayed in the table). However, and in line with the analysis of Section 3.4, the rejection of the null hypothesis for cross trading at higher lags is not so strong; on the other hand, the further lags included for parallel trading are almost always not significant. The two lags of D_k included in the variance of the heteroskedastic ordered probit do not exhibit a homogeneous pattern in the whole sample, and a general conclusion on the significance of these variables cannot be drawn. Conversely, and more interestingly, the time and the seasonality components are generally significant for all the stocks. Time always exhibits a positive sign, which is in line with the results discussed in HLM (1992). This implies a positive clock-time effect for the conditional variance of D_k^*, with time elapsing that is associated with an increasing variance. As to seasonality, a LR test to jointly exclude the terms of the Fourier series is usually rejected, which confirms the presence of some form of periodicity for the intraday volatility.

3.8 Informative Content of Observed Market Variables

Section 3.5 discussed the impact of informed and uninformed traders on market transactions, and it shows that a significant effect exists only when the type of traders is different on the two sides of the market. Since traders' identities at Euronext are not available to operators, neither in real time nor with delay, it is worth investigating more accurately the possible causes of this result. In this section, we examine a wide set of LOB variables and we explore their informativeness to detect informed-based trading.

3.8.1 The bivariate probit model for the identity of traders

In the case of anonymous markets, the investors cannot build their trading strategies by replicating the behaviour of informed operators who are not identifiable; however, they can extract the relevant information behind a specific trading pattern. To form a guess on traders' identities, we consider a comprehensive list of variables including inter-trade durations, intraday trading patterns, market prices, and transaction volumes. Easley and O'Hara (1992), Biais et al. (1995), Gourieroux et al. (1999), and Foucault et al. (2007)

among others, have highlighted the role of volume, time, stock price, spread, volatility, and daily periodicity to identify the origin of transactions. All these variables represent public information, as long as they are visible to market members, or easily recoverable from the LOB. Some of these variables act as driving elements for automated trading algorithms like the volume-weighted average price (VWAP) or the time-weighted average price (TWAP), see Bialkowski et al. (2008) or Brownlees et al. (2010). To examine how these variables can convey information on the traders' identity, we adopt the following bivariate probit model:

$$\begin{aligned}
Pr(D_{b_k} = 1|\mathbf{X}_k) &= \Phi(\mathbf{X}_k'\beta_b + \epsilon_{1k}) \\
Pr(D_{s_k} = 1|\mathbf{X}_k) &= \Phi(\mathbf{X}_k'\beta_s + \epsilon_{2k}) \\
Cov(\epsilon_{1k}, \epsilon_{2k}) &= \rho
\end{aligned}$$

where X_k includes the set of explanatory variables, β_b and β_s represent the parameter vectors, ϵ_{1k} and ϵ_{2k} are Gaussian noises with correlation coefficient ρ, while D_{b_k} and D_{s_k} are two dummy variables that are equal to one when the trader is informed on the buy side and on the sell side, respectively. According to the classification introduced in Section 3.3, $D_{b_k} = 1$ when the buyer is coded with '2' or '7', while $D_{b_k} = 0$ when the buyer is coded with '1'; the same sorting immediately applies to the case of sellers as well. As explanatory variables we consider:

- A set of time indicators which distinguish some specific moments of the continuous trading session: D_{open}, D_{lunch}, D_{SP}, and D_{clos}. The variable D_{open} identifies a transaction that occurs between 9.00 A.M. and 9.30 A.M., D_{lunch} between 00.30 P.M. and 1.30 P.M., D_{SP} between 3.30 P.M. and 4.00 P.M. and D_{clos} between 5.00 P.M. and 5.30 P.M.. These time dummies are associated with some critical periods of the trading session and they have been frequently used to test different microstructure hypotheses, see Dufour and Engle (2000) or Lo and Sapp (2010). The opening and the closing 30 minutes (D_{open}, D_{clos}) are usually characterized by high price volatility and frequent transactions, while the opposite generally holds for the lunch time (D_{lunch}). D_{SP} individuates the trades occurring within the first 30 minutes from the opening of the NYSE, when trading from institutional investors is more frequent. We also include a variable δ_k, going from zero to one, as discussed in Section 3.4 for daily seasonality; this regressor is intended to measure the time evolution throughout the continuous auction.

- The time between consecutive transactions, Δt_k. In their seminal work, Easley and O'Hara (1992) analyzed the informativeness of trade durations and conclude that a lower trading frequency is usually associated with a lower presence of informed investors in the market. This is explained by the fact that informed traders are impatient to profit from their informational privilege; thus, a long trade duration is likely to be associated with no private information in the market. This issue has been widely tested (e.g. Dufour and Engle, 2000 or Manganelli, 2005) and we expect Δt_k to exhibit a negative impact on the probability of informed trading.

- Transaction volume, wherein informativeness is expressed through $Volume$ and D_{big}. A large order represents a sort of bet for traders, as it exposes the investors to a higher potential loss; therefore, the volume is perceived by the market as a reliable support for a particular trading strategy. It is commonly accepted in the literature that institutional traders post larger orders compared to retail investors (Easley and O'Hara, 1987). This stylized fact is confirmed in our sample, and the average volume per category of investor is omitted only to save space. Several reasons are behind this observation, e.g. institutional traders want to extensively profit from their private information or they enter the market with larger orders, as for the investment funds (Biais et al., 1995). Generally speaking, this means that transactions displaying a sizeable volume are more likely executed by institutional agents, and we expect a positive sign for both regressors. The variable *Volume* captures the impact of quantity by including the number of shares per transaction, while the dummy variable D_{big} explicitly identifies the sizeable transactions, as it is equal to 1 if a transaction displays a volume larger than the average of the previous 15 minutes.

- Squared variation, SV. It is related to the definition of realized variance, which represents one of the standard volatility measure with high-frequency data, see Andersen et al. (2003). To define SV, we partition the trading session into intervals of 15 minutes each, wherein SV is computed as the squared log difference between transaction prices at the beginning and the end of each interval. SV is included with one lag to avoid simultaneity bias, and it is multiplied by 100 to obtain a comparable estimate in terms of magnitude. Clearly, the purpose of SV is far from representing a perfect measure of the intraday volatility; however, we adopt SV to investigate the relationship between large price variations and the presence of informed-based trading. Several papers has examined this topic, see Daigler and Wiley (1999), Ahn et al. (2001) or Manganelli (2005). Given that informed-based trading tends to be highly clustered, if the price variation is mainly driven by the liquidity traders, we expect a negative sign for SV, while the opposite holds when the price variation is generated by a private information signal.

- The variable that identifies the side of the market which initiates a transaction $Init$, defined exactly as in Section 3.4. The previous empirical research has emphasized the role of informed traders as initiators of the market transactions, see Chakravarty (2001). We expect $Init$ to display an opposite effect on D_{b_k} and D_{s_k}, according to the side of the market that has generated the transaction. Indeed, when the transaction is buyer initiated, it should more likely have originated from an informed buyer and the symmetrical hypothesis clearly holds for the sell side.

From the whole set of regressors, we distinguish *time*-related variables (D_{open}, D_{lunch}, D_{SP}, D_{clos}, δ_k, and Δt_k), *volume*-related variables ($Volume$ and D_{big}), and, in a broader sense, *price*-related variables (SV and $Init$). To proceed in the estimation, we split the time series of each stock into two subsamples, in order to produce in-sample and out-of-sample estimates; without loss of generality, the first 80% of the observations is dedicated to achieve in-sample results. We evaluate the forecasting performance of our model with the quadratic probability score (QPS) defined by Diebold and Rudebusch (1989):

$$QPS = 1/T \sum_{k=1}^{T} 2(P_k - D_k)^2 \tag{3.7}$$

where P_k represents the bivariate probit probability forecast, and D_k is the corresponding observed realization. The QPS ranges from 0 to 2, where 0 stands for the perfect model prediction; this measure has been applied for bivariate probit model by Nyberg (2009). In our context, what actually matters is detecting the presence of institutional trading on at least one of the two market sides. This can be done by computing the following marginal conditional probabilities:

$$P_{b_k} = P_{11_k} + P_{10_k}$$
$$P_{s_k} = P_{11_k} + P_{01_k}$$

where P_{b_k} measures the marginal conditional probability that the buyer is an informed traders, while P_{s_k} measures the same statistics for the sell side. More precisely, P_{b_k} is the sum of two probabilities: P_{11_k} that expresses the likelihood of observing an informed agent on both market sides, and P_{10_k} that considers the case of an informed buyer and an uninformed seller. An analogous definition straightforwardly follows for P_{s_k}. We use the two probability forecasts P_{b_k} and P_{s_k} in Equation (3.7) to assess the accuracy of the bivariate probit estimates.

3.8.2 The bivariate probit model: results

Table 3.12 provides the parameter estimates of the bivariate probit model for Bouygues and BNP. A discussion of the results limited to these two stocks would not offer an

Table 3.12: Bivariate probit estimates for Bouygues and BNP, with p-values in parentheses. The bottom lines display goodness-of-fit statistics.

	Bouygues		BNP	
	Buy	Sell	Buy	Sell
D_{open}	-9.94e-02	-1.45e-01	-6.37e-02	7.60e-02
	(0.00)	(0.00)	(0.00)	(0.00)
D_{lunch}	-3.93e-02	-4.28e-02	4.85e-02	2.06e-02
	(0.00)	(0.00)	(0.00)	(0.00)
D_{SP}	-1.37e-02	6.52e-02	-3.75e-02	-2.78e-04
	(0.21)	(0.00)	(0.00)	(0.97)
D_{clos}	-6.89e-02	1.05e-01	4.57e-04	-7.51e-02
	(0.00)	(0.00)	(0.95)	(0.00)
δ_k	3.72e-06	3.27e-06	7.49e-06	1.08e-05
	(0.00)	(0.00)	(0.00)	(0.00)
Δt_k	-1.69e-03	-1.58e-03	-6.68e-03	-5.70e-03
	(0.00)	(0.00)	(0.00)	(0.00)
Volume	3.54e-06	2.17e-05	-1.73e-05	-3.59e-05
	(0.64)	(0.01)	(0.00)	(0.00)
D_{big}	1.14e-01	8.59e-02	7.64e-02	5.77e-02
	(0.00)	(0.00)	(0.00)	(0.00)
Init	1.49e-01	-9.28e-02	4.33e-02	-4.84e-02
	(0.00)	(0.00)	(0.00)	(0.00)
SV	-7.84e+00	-1.81e-01	-2.05e+00	-1.72e+00
	(0.00)	(0.61)	(0.00)	(0.00)
Constant	9.36e-02	1.85e-01	-7.09e-03	6.53e-02
	(0.00)	(0.00)	(0.11)	(0.00)
QPS_{in}	0.48	0.47	0.49	0.48
QPS_{out}	0.49	0.51	0.49	0.49

exhaustive analysis, so the following examination is also addressed on the basis of Table 3.13 where the results for the whole sample (39 stock series) are summarized. Table 3.13 splits the whole-sample results for both sides of the market and it provides the percentages of significant and non-significant estimates; a positive sign is supportive of informed-based trading, while the opposite holds for a negative estimate.

Except for D_{open}, it is quite difficult to draw a general conclusion on the impact of the time dummy variables. Indeed, D_{open} is the only time dummy that exhibits a fairly homogeneous result, with a negative and statistically significant coefficient for around two-thirds of the series on the buy side, and almost half of the series on the sell side.

Table 3.13: Bivariate probit estimates for the whole sample. The first column exhibits the percentage of negative and significant estimates, the central column shows the percentage of non-significant coefficients, while the last one reports the percentage of positive and significant estimates.

	Buy side		
	Negative Significant	Non-significant	Positive Significant
D_{open}	66.67	28.21	5.13
D_{lunch}	28.21	28.21	43.59
D_{sp}	30.77	53.85	15.38
D_{clos}	38.46	28.21	33.33
δ_k	5.13	0.00	94.87
Δt_k	74.36	23.08	2.56
Volume	33.33	38.46	28.21
D_{big}	10.26	12.82	76.92
Init	0.00	0.00	100.00
SV	76.92	15.38	7.69

	Sell side		
	Negative Significant	Non-significant	Positive Significant
D_{open}	48.72	30.77	20.51
D_{lunch}	25.64	41.03	33.33
D_{sp}	15.38	51.28	33.33
D_{clos}	35.90	30.77	33.33
δ_k	2.56	5.13	92.31
Δt_k	84.62	15.38	0.00
Volume	38.46	33.33	28.21
D_{big}	7.69	7.69	84.62
Init	100.00	0.00	0.00
SV	74.36	17.95	7.69

On the other hand, the positive estimates are extremely marginal for the buy side and around 20% for the sell side, with both sides displaying almost one-third of non-significant coefficients. On the whole, these results suggest a strong occurrence of transactions executed on behalf of retail investors during the first 30 minutes of the continuous auction, which is in line with Biais et al. (1995) or Gourieroux et al. (1999). Biais et al. (1995) showed that smaller trades usually occur during the morning, while larger trades are more frequent in the late afternoon. As an explanation, they suggest that financial intermediaries prefer to trade retail orders at the beginning of the day, postponing their own trading in the afternoon because of price discovery. This is particularly true at the end of the continous auction, when the necessity to close open positions is more urgent. However, we only find a partial confirmation of the last point, as the effect of D_{clos} is split almost equally between informed, uninformed and non-significant estimates, and the likelihood of observing a higher concentration of informed traders in the last 30 minutes is unclear. The estimates of D_{lunch} for the buy side are consistent with a large fraction of institutional investors on the market; the same effect is less apparent for the sell side, though it still reveals a higher presence of informed agents. The general decrease in the trading frequency observed at lunch time seems counterbalanced by a higher number of institutional transactions; in any case, the fraction of non-significant estimates is quite high, at approximately one-third of the sample. The findings for D_{SP} are different for the two sides of the market. For D_{s_k}, transactions occurring immediately after the opening of the US Exchange are coherent with the presence of institutional investors for one-third of the sample, with a large partition of the stocks exhibiting non-significant estimates. On the contrary, in the case of D_{b_k}, transactions executed from 3.30 P.M. and 4.00 P.M. are more likely to be implemented by uninformed investors. Generally, the high fraction of non-significant estimates for D_{lunch} and D_{SP} actually prevents us from drawing any specific conclusion for both regressors. On the other hand, the result relative to δ_k is noticeable: in more than 90% of the sample, the likelihood of informed-based trading increases as time elapses during the continuous auction. This result holds for both D_{b_k} and D_{s_k}, and it perfectly matches Biais et al. (1995) who emphasize the increase of institutional trading when moving towards the closing time. Although D_{clos} shows that these findings are not confirmed for the last 30 minutes of the continuous session, we do not deem this outcome as self-contradictory: indeed, institutional investors may just be willing to reduce their trading in periods of high price volatility, such as the one very close to the end of the day.

The results concerning trade durations strongly support the thesis of Easley and

O'Hara (1992) and our estimates report that the increase of inter-trade durations is generally associated with a lower presence of informed agents. This result is evident from Table 3.13, where a negative sign is found for almost 75% of the stocks on the buy side and more than 80% on the sell side. These conclusions also hold when an additional lag of Δt_k is included in the bivariate probit specification (not reported), though the strength of the effect is generally weaker.

Volume and D_{big} are the two variables deputed to evaluate the volume informativeness, but the inspection of Table 3.13 delivers an equivocal result for the role of volume as a signal of informed-based trading. First of all, the actual transaction volume exhibits a high fraction of non-significant estimates, larger than 30% of the series for both sides of the market. Moreover, it is hard to clearly state if the actual volume has a positive or a negative effect, though the negative estimates are more frequent, especially for the sell side. On the contrary, the behaviour of D_{big} is quite homogeneous in the whole sample. For the 76.92% of the stocks on the buy side and 84.62% on the sell side, the probability of informed trading increases when the market presents transactions with a larger-than-average volume. These findings may appear misleading at first glance, but they do not seem so unreasonable. The estimates show that the dimension of market transactions *per se* is not very much informative. Nevertheless, when the transaction size proves to be larger than its recent average, the volume becomes highly revealing of the traders' identities, in line with the theoretical model of Easley and O'Hara (1987).

As for price-related variables, Table 3.12 shows that SV has a negative influence on the probability of trading by informed agents. The result is also validated by Table 3.13 where, more than the 70% of the series displays a negative estimate on both market sides. The negative sign of SV shows that informed-based trading is unlikely to be associated with periods of large price variations, as it was already mentioned in the previous analysis of the time dummies D_{open} and D_{clos}. Rather, by following the distinction in Dufour and Engle (2000) or Wong et al. (2008), our findings indicate that the large price variations are mostly generated by liquidity traders.

With respect to $Init$, Table 3.12 displays a positive coefficient for D_{b_k} and a negative coefficient for D_{s_k}; this result spreads over all the stock series as it is evident from Table 3.13. For buyer-initiated transactions, the positive coefficient for D_{b_k} indicates a higher probability of informed trading on the buy side. Conversely, the opposite interpretation holds for D_{s_k}, and in the case of buyer-initiated transactions, it is unlikely that the seller is an informed agent. Besides, these results highlight the role of institutional operators as trade initiators and they are consistent with the assumption that retail investors mainly act as liquidity traders.

Finally, at the bottom of Table 3.12, the QPS values for in- and out-of-sample estimates are provided. The two values are around 0.50, suggesting a quite appropriate fit. These numbers refer only to Bouygues and BNP, but they extend similarly to the whole data set, where we reasonably find a better prediction for the in-sample estimates.

3.9 Conclusions

Using high-frequency data from Euronext Paris, we analyse the impact of informed and uninformed traders on market transactions. We adapt the framework described in HLM (1992) to verify the presence of a trader effect at the transaction level. Our results show that institutional investors affect market prices when they are matched with retail investors. Informed buyers transmit a positive pressure to the market when they trade with an uninformed seller, while the opposite holds when an institutional seller trades with an uniformed buyer. On the other hand, no significant effect is encountered when the traders coincide on the two sides of the market. Our findings are robust to alternative model specifications, and generally extend to the whole data set; nevertheless, they are not fully expected, since traders' identities are concealed at Euronext Paris. Therefore, in the last part of the paper, we also examine the informativeness of a wide set of market variables, and we explore their usefulness to infer the identity of investors. We find informed trading to be more likely to occur as time elapses during the continuous auction and in periods of high-frequency of transactions. Conversely, informed trading is usually absent at the start of the trading day and during periods of large price variations. Finally, informed traders generally operate with larger-than-the-average volumes and they are found to act as initiators of the market transactions.

Bibliography

Admati, A.R., and Pfleiderer, P. (1988),"A Theory of Intraday Patterns: Volume and Price Variability," *The Review of Financial Studies*, 1, 3-40.

Alexander, G.J. and Peterson, M.A. (2007), "An Analysis of Trade-Size Clustering and its Relation to Stealth Trading," *Journal of Financial Economics*, 84, 435-471.

Andersen T.G., Bollerslev T., Diebold F.X., and Labys P. (2003), "Modeling and forecasting realized volatility", *Econometrica*, 71, 579-625.

Barclay, M., and Warner, J. (1993), "Stealth Trading and Volatility: which Trades Move Prices?," *Journal of Financial Economics*, 34, 281-305.

Biais, B., Glosten, L., and Spatt, C. (2005), "Market Microstructure: a Survey of Microfundations, Empirical Results and Policy Implications," *Journal of Financial Markets*, 8, 217-264.

Biais, B., Hillion, P., and Spatt, C. (1995), "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *The Journal of Finance*, 50, 1655-1689.

Bialkowski, J., Darolles, S. and Le Föl, G. (2008), "Improving VWAP Strategies: a Dynamic Volume Approach," *Journal of Banking and Finance*, 32, 1709-1722.

Brownlees, C.T., Cipollini, F. and Gallo, G.M. (2010), "Intra-Daily Volume Modeling and Prediction for Algorithmic Trading," *working paper*.

Chakravarty, S. (2001), "Stealth Trading: which Traders' Trades Move Stock Prices?," *Journal of Financial Economics*, 61, 289-307.

Daigler, R.T., and Wiley, M.K. (1999), "The Impact of Trader Type on the Futures Volatility-Volume Relation" *The Journal of Finance*, 54, 2297-2316.

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

Diebold, F.X., and Rudebusch, G.D. (1989), "Scoring the Leading Indicators," *The Journal of Business*, 62, 369-391.

Dufour, A., and Engle, R.F. (2000), "Time and the Price Impact of a Trade," *The Journal of Finance*, 55, 2467-2498.

Easley, D., Hvidkjaer, S., and O'Hara, M. (2002), "Is Information Risk a Determinant of Asset Returns?,", *The Journal of Finance*, 57, 2185-2221.

Easley, D., Kiefer, N.M. and O'Hara, M. (1997), "The Information Content of the Trading Process," *Journal of Empirical Finance*, 4, 159-186.

Easley, D., Kiefer, N.M., O'Hara, M. and Paperman, J.B. (1996), "Liquidity, Information and Infrequently Traded Stocks," *The Journal of Finance*, 51, 1405-1436.

Easley, D., and O'Hara, M. (1987), "Price, Trade and Information in Securities Markets," *Journal of Financial Economics*, 19, 69-90.

Easley, D., and O'Hara, M. (1992), "Time and the Process of Security Price Adjustment," *The Journal of Finance*, 47, 577-605.

Easley, D., and O'Hara, M. (1997), "High-Frequency Data in Financial Markets: Issues and Applications," *Journal of Empirical Finance*, 4, 73-114.

Easley, D., Engle, R.F., O'Hara, M., and Wu, L. (2008), "Time-Varying Arrival Rates of Informed and Uninformed Trades" *Journal of Financial Econometrics*, 6, 171-207.

Ellis, K., Michael, R., and O'Hara, M. (2000), "The Accuracy of Trade Classification Rules: Evidence from Nasdaq," *Journal of Financial and Quantitative Analysis*, 35, 529-551.

Engle, R.F., and Patton, A.J. (2004), "Impact of Trades in an Error-Correction Model of Quote Prices," *Journal of Financial Markets*, 7, 1-25.

Foucault, T., Moinas, S., and Theissen, E. (2007), " Does Anonymity Matter in Electronic Limit Order Markets?," *The Review of Financial Studies*, 5, 1707-1747.

Glosten, L., and Harris, E. (1988), "Estimating the ComponentS of the Bid-Ask Spread," *Journal of Financial Economics*, 21, 123-142.

Glosten, L., and Milgrom, J. (1985), "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71-100.

Goodhart, C.A.E., and O'Hara,M. (1997), "High-Frequency Data in Financial Markets: Issues and Applications," *Journal of Empirical Finance*, 4, 73-114.

Gourieroux C., Jasiak, J., and Le Fol, G. (1999), "Intra-Day Market Activity," *Journal of Financial Markets*, 2, 193-226.

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987) , "Generalized Residuals," *Journal of Econometrics*, 34, 1-52.

Hasbrouck, J. (1991), "The Summary Informativeness of Stock Trades: an Econometric Analysis," *The Review of Financial Studies*, 4, 571-595.

Hasbrouck, J. (2007), "Empirical Market Microstructure," *Oxford University Press.*

Hausman, J., Lo, A. and MacKinlay, A.C. (1992), "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319-379.

Kyle, A. (1985), "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315-35.

Lee, C.M.C., and Ready, M.J. (1991), "Inferring Trade Direction from Intraday Data," *The Journal of Finance*, 46, 733-746.

Liesenfeld, R., Nolte, I. and Pohlmeier, W. (2006), "Modelling Financial Transaction Price Movements: a Dynamic Integer Count Data Model," *Empirical Economics*, 30, 795-825.

Lo, A., MacKinlay, C. and Zhang, J. (2001), "Econometric Models of Limit-Order Executions," *Journal of Financial Economics*, 65, 31-71.

Manganelli, S. (2005), "Duration, Volume and Volatility Impact of Trades," *Journal of Financial Markets*, 8, 377-399.

Nyberg, H. (2009), "A Bivariate Autoregressive Probit Model: Predicting U.S. Business Cycle and Growth Rate Cycle Recessions", *working paper*.

Odders-White, E.R. (2000), "On the Occurrence and Consequences of Innacurate Trade Classification," *Journal of Financial Markets*, 3, 259-286.

Roll, R. (1984), "A Simple Implicit Measure of the Bid-Ask Spread in an Efficient Market," *The Journal of Finance*, 39, 1127-39.

Williams, R. (2009), "Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients across Groups," *Sociological Methods & Research*, 37(4), 531-559.

Chapter 4

Traders and Time: Who Moves the Market?

Fabrizio Ferriani

JEL codes: C10, C41, G10, G14

Keywords: Market microstructure, Financial durations, Log-ACD model, Euronext Paris, Informed traders, Liquidity provider

4.1 Introduction

The microstructure of financial markets offers several areas of investigation for researchers and regulators. One of the most promising fields concerns the analysis of financial durations, which has recently benefited from an increasing quality in data recording. With the expression 'financial durations', we are referring to a time interval between two consecutive market events. As an example, trade durations measure the time between two subsequent transactions, while volume durations quantify the time necessary to exchange a specific amount of shares. Theoretically speaking, the interest on time and its informational implications was firstly recognized by Diamond and Verecchia (1987) and Easley and O'Hara (1992). Diamond and Verecchia (1987) propose a model that excludes the possibility of short selling by informed agents, such that the absence of trading is associated with the existence of bad news. Conversely, in the Easley and O'Hara (1992) setting, the traders are not subjected to sell constraints and a long duration is interpreted as a lack of information in the market.

In the empirical analysis of financial durations, a highly relevant contribution is advanced by Engle and Russell (1998) who introduce the autoregressive conditional duration (ACD) model. They examine the IBM trade durations and observe a clustering pattern similar to the one occurring in price volatilities. Therefore, they suggest the study of trade durations with a dynamic approach, which essentially imitates the original ARCH model proposed by Engle (1982). From that moment on, the empirical research on financial durations has been widely extended by following two main, and sometimes overlapping, directions. A first line of research concentrates on the definition of a more flexible specification with respect to the basic ACD model, which proves to be not particularly suited to fit market data. This vein includes the logarithmic ACD model (Bauwens and Giot, 2000) which breaks the linearity between the current duration and its previous (conditional) lags or the stochastic conditional duration model (Bauwens and Veredas, 2004) which assumes a non-deterministic structure for the conditional duration; a further contribution is also given by the threshold ACD model (Zhang et al., 2001) which introduces a mechanism of switching regimes in the modelling of financial durations. Along this first line of research, some papers also seek for alternative distributions to the original exponential and Weibull densities that generally exhibit a clear misspecification in the diagnostic tests. This is the case of Lunde (1999) and Bauwens (2006) who consider the Generalized Gamma distribution, or Grammig and Maurer (2000) who adapt the ACD model to the Burr distribution. On the other hand, a second line of empirical investigation dedicates to the testing of various microstructure hypotheses. In this case, the usual strategy consists of augmenting one of the previous models with a set of additional variables, such as transaction volume, bid-ask spread, trading intensity, or some volatility proxy. Most of the time, this approach is applied to the so-called aggregated durations (e.g. volume or price durations) because of their relationship with some market features; some examples may be found in Bauwens and Giot (2000), Bauwens and Giot (2003), Bauwens and Veredas (2004), and Wong et al. (2008).

Our analysis extends the last vein of research and it explores a new microstructure hypothesis. Indeed, the main and original contribution of this paper is to study the impact of traders on time. This is made possible by the information available in our data set, where we can identify the different categories of traders operating in the market, in particular the so-called informed traders and the liquidity provider. The motivation behind their presence in the market is obviously different, as the former want to exploit private information, while the latter trades because of contractual constraints. Nevertheless, we expect that the overall market activity is fostered by the increasing presence of these traders. Our assumption is tested and confirmed on three distinct definitions of financial durations. The findings prove to be robust across alternative distributions, as well as when we augment the basic model with additional microstructure variables.

The paper is organized as follows. Section 4.2 reviews the fundamental econometric models which are traditionally used for the analysis of financial durations and it examines the distributional assumptions. Section 4.3 describes the data, while Section 4.4 introduces the three types of financial durations and it provides some descriptive statistics. Section 4.5 presents the estimation strategy and illustrates the empirical specification. Section 4.6 discusses the main results, while Section 4.7 presents some robustness tests and estimation diagnostics. Finally, Section 4.8 offers our conclusions.

4.2 Econometric models

4.2.1 The basic setting

Without loss of generality, we define a duration $d_i = t_i - t_{i-1}$ as the time interval between two consecutive market events, whether they be trades or not. Engle and Russell (1998) propose a model where the serial dependence in trade durations is captured through the conditional expectation $\Psi_i = E(d_i|\Omega_{i-1})$, with Ω_{i-1} representing the filtration at time $i-1$. Their framework is based on two equations; the first one specifies a multiplicative error term structure for the observed duration d_i:

$$d_i = \Psi_i \epsilon_i \tag{4.1}$$

where ϵ_i are positive IID random variables, with $E(\epsilon_i) = 1$ and $Var(\epsilon_i) = \sigma^2$. The hypothesis about $E(\epsilon_i)$ is consistent with $\Psi_i = E(d_i|\Omega_{i-1})$, although a simple reparametrization also allows for $E(\epsilon_i) = \mu \neq 1$. The second equation is used to explain how the time dependence is transmitted through the conditional duration Ψ_i. Engle and Russell (1998) suggest a linear autoregressive structure for the conditional mean function:

$$\Psi_i = \omega + \alpha d_{i-1} + \beta \Psi_{i-1} \tag{4.2}$$

which requires $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ as constraints to ensure the positivity of Ψ_i. Equation (4.2) refers to the case of an ACD (1,1) model, but it can be straightforwardly extended to include additional lags of d_i and Ψ_i.

Engle and Russell (1998) originally adopt the exponential and the Weibull distribution to estimate their ACD model. The exponential distribution delivers QML estimates under the assumption of correct specification of the conditional mean Ψ_i. The Weibull distribution introduces some form of flexibility, and it allows a non-constant hazard function according to the unconstrained shape parameter. Once the innovation density for ϵ_i is defined, the full vector of model parameters may be estimated through the loglikelihood function:

$$l(\theta_1; \theta_2) = \sum_{i=1}^{n} \left[\ln f_\epsilon \left(\frac{d_i}{\Psi_i(\theta_2)}; \theta_1 \right) - \ln \Psi_i(\theta_2) \right] \tag{4.3}$$

where f_ϵ represents the error term distribution, θ_1 collects its associated parameters, and θ_2 groups the parameters of the conditional duration Ψ_i.

4.2.2 Extended frameworks

The original setting proposed by Engle and Russell (1998) subsequently evolves to meet the multiple requirements coming from the market microstructure research. A first extension, which we also adopt in this paper, is represented by the Log-ACD model described in Bauwens and Giot (2000). The ACD model could raise some concerns if equation (4.2) is augmented by a vector of variables z_i to study their impact on financial durations. Actually, such an issue is problematic whenever the variables included in z_i exhibit a negative coefficient; in this case, the conditional duration Ψ_i could turn out to be negative, which is not acceptable for a strictly positive variable. The easiest solution is to impose some positivity constraints on the z_i coefficients; however, this is clearly conflicting with the purpose of testing some microstructure hypotheses. As a way to solve this problem, Bauwens and Giot (2000) suggest a specification which reminds of Nelson (1991) EGARCH model. More precisely, they rewrite equation (4.1) as:

$$d_i = \exp(\psi_i)\epsilon_i \tag{4.4}$$

where $\psi_i = \ln \Psi_i$ represents the logarithm of the conditional duration. A thorough review of the Log-ACD model, as well as the definition of its moments and stationarity conditions, is out of the scope of this research, and we refer the interested reader to Bauwens and Giot (2000), Bauwens and Giot (2001), and Bauwens et al. (2003). Here, we only outline the two possible specifications for $\ln \Psi_i$ described in Bauwens and Giot (2000):

$$\begin{aligned} \ln \Psi_i &= \omega + \alpha \ln \epsilon_{i-1} + \beta \ln \Psi_{i-1} \qquad (4.5) \\ &= \omega + \alpha \ln d_{i-1} + (\beta - \alpha) \ln \Psi_{i-1} \end{aligned}$$

and

$$\begin{aligned} \ln \Psi_i &= \omega + \alpha \epsilon_{i-1} + \beta \Psi_{i-1} \\ &= \omega + \alpha (d_{i-1}/\Psi_{i-1}) + \beta \ln \Psi_{i-1} \end{aligned} \tag{4.6}$$

where the first one is traditionally labelled as Log-ACD$_1$, while the second one is referred to as Log-ACD$_2$. Again, Equations (4.5) and (4.6) show the basic dynamic setting, which can be easily extended to include additional lags of d_i or ψ_i. The two specifications deliver similar estimates and do not require any sign restrictions to ensure positivity, making the Log-ACD model particularly suited for testing the impact of microstructure variables[1].

As was anticipated in the Introduction, a further line of improvement with respect to Engle and Russell (1998) is regarding the choice of the innovation density for ϵ_i. Indeed, the exponential and the Weibull distributions prove to be inappropriate to replicate the patterns observed for financial duration series. As an example, Bauwens et al. (2004) show that the exponential and the Weibull distributions tend to overestimate small durations and underestimate very small durations. This essentially happens because these distributions present a too much restrictive hazard function that is not satisfactorily reconcilable with market data. The literature generally concentrates on just a few options as possible substitutes for the exponential and the Weibull distributions. In this paper, we provide estimates based on the Burr and the Generalized Gamma distributions which both encompass the exponential and the Weibull specifications as special cases. The Burr distribution is defined as:

$$f_B(\epsilon_t) = \frac{\gamma}{c} \left(\frac{\epsilon_t}{c} \right)^{\gamma - 1} \left[1 + \lambda \left(\frac{\epsilon_t}{c} \right)^{\gamma} \right]^{-(1+\lambda^{-1})}$$

[1]For the testing of microstructure hypotheses, there exist alternative specifications which could be used instead of the Log-ACD model. Nevertheless, as shown in Bauwens et al. (2004), a more complex and time-demanding framework does not generally perform better than a Log-ACD model estimated with a flexible distribution. Coherently, we adopt the Log-ACD specification and we complement our estimates with a comprehensive section of robustness tests and diagnostics.

where $\gamma > 0$, $\lambda > 0$, and $c > 0$ are parameters. By setting

$$c = \lambda^{(1+\lambda^{-1})} \frac{\Gamma(1+\lambda^{-1})}{[\Gamma(1+\gamma^{-1})\Gamma(\lambda^{-1}-\gamma^{-1})]}$$

the mean is equal to one and the density reduces to a two parameter distribution, $f_B(\gamma, \lambda)$. The Generalized Gamma distribution has a density function:

$$f_{GG}(\epsilon_t) = \frac{\gamma}{c^{\nu\gamma}\Gamma(\nu)} \epsilon_t^{\nu\gamma-1} \exp\left[-\left(\frac{\epsilon_t}{c}\right)^{\gamma}\right]$$

where $\nu > 0$, $\gamma > 0$ and $c > 0$ are parameters, while $\Gamma(\nu)$ is the gamma function. If c is set equal to $\frac{\Gamma(\nu)}{\Gamma(\nu+\gamma^{-1})}$, then $E(\epsilon_t) = 1$ and the Generalized Gamma may also be written as a two parameter distribution, $f_{GG}(\nu, \gamma)$. Some previous contributions (e.g. Bauwens et al., 2004) highlighted that these two densities increase the model flexibility, as they are able to break the one-to-one correspondence between the durations and the properties of the hazard function.

4.3 Data

The data set for the empirical analysis is supplied by Eurofidai and it spans over a two-month period, from 1 September 2009 to 30 October 2009. It provides exhaustive information for five stocks belonging to the CAC40 index, namely: Alstom, Axa, Crédit Agricole, Sanofi-Aventis, and Schneider. The data set reports a wide set of variables concerning market transactions as well as some valuable information to qualify the market participants. More precisely, the data set presents a variable to distinguish the categories of traders operating in the market, according to the same codes discussed in Ferriani (2012):

- '1' is the code that refers to transactions executed on behalf of retail investors;
- '2' is the code that refers to transactions executed by operators authorized to trade

in the Paris Bourse (banks or other financial intermediaries, called 'Sociétés de Bourse');

- '6' is the code that refers to transactions executed by 'fournisseurs de la liquidité', i.e. agents with liquidity duties;

- '7' is the code that refers to transactions executed by another kind of financial intermediaries called 'Filiales de la Société de Bourse'. They are financial institutions similar to the ones classified with '2'.

For each transaction, the data set displays a code for both sides of the market, such that the buyer and the seller can either belong to the same category of traders or they can be different. To distinguish among traders, we adopt a classification which is standard in the empirical microstructure literature (see Chakravarty, 2001). The observations coded with '1' are associated with transactions executed on behalf of retail investors, while observations classified with '2' or '7' refer to transactions executed by financial intermediaries in their own interest. Conversely to Ferriani (2012), the liquidity provider exhibits an active role in this data set, and a considerable amount of transactions (around 15-20%) is marked with code '6', either on the buy or on the sell side. Indeed, though Euronext Paris is essentially an order-driven market, the Stock Exchange admits the presence of a liquidity provider who facilitates the trading activity by posting trade proposals on both market sides.

Finally, for the purpose of this research, we have to mention the reform which introduced anonimity at Euronext Paris in 2001. Under the new market regime, the traders are no longer able to view, not even with delay, the operator who is trading or the category he belongs to. This aspect is also discussed in Foucault et al. (2007) who analyze the effects of the switch from a fully disclosed market to a regime with hidden identities. As a matter of fact, anonymity in financial markets is not a secondary issue, as it excludes a 'direct' imitation effect among market participants. More precisely, the presence of hidden identities rules out the strategies that exactly replicate the behaviour of informed traders, as the latter are not unambiguously identifiable through an ID code[2].

4.4 Financial durations

4.4.1 Definitions

We recover three definitions of financial durations from the data set; in the following, unless otherwise specified, we adopt d_i to refer to any of them. First of all, we consider trade durations, which measure the time elapsed between two consecutive transactions. Transaction time is registered at microsecond-level precision, but we choose to round off microseconds at the third decimal digit. This leaves the main results unaffected, meanwhile reducing the risk that extremely small durations prevent the optimization algorithm from finding a solution. The microseconds also influence the occurrence of null durations, which take place when a single large order is almost simultaneously matched with multiple orders on the opposite market side. This splitting generates a sequence of transactions recorded at the same time second, which conflicts with the positivity required to estimate duration models. The literature proposes different approaches to deal with null durations, though a prevailing strategy is not clear. As an example, in Dufour and Engle (2000a) zero durations are simply discarded, while Dufour and Engle (2000b) indistinctly add one second to each observation. Bauwens (2006) similarly includes an additional second, but only to simultaneous transactions. Conversely, Veredas et al. (2002) suggest a procedure to equally distribute time across consecutive null durations. The availability of observations registered at microsecond-level precision sensibly downsizes the problem of null durations; nevertheless, the rounding-off introduced to estimate our model still leaves some simultaneous observations. Therefore, we choose to group simultaneous transactions that presumably belong to the same unique trade according to these

[2]Clearly, several microstructure variables may be helpful to detect the categories of investors, though most of the times these variables are valuable for just shaping a hypothesis about the trader identity. Conversely, here we are considering a more revealing information, i.e. the exact identification of the trader or of the category he belongs to.

steps:

- First, the time stamp expressed as 'hh:mm:ss' has to be the same, i.e. observations must be simultaneous at least with respect to seconds;
- Second, simultaneous observations are cumulated only if they are generated by the same side of the market[3];
- Finally, simultaneous observations are cumulated only if they display a uniform trader code (1, 2, 6 or 7) on the side of the market which initiates the transaction.

At the end of this procedure, we have grouped all the simultaneous observations that come from the same side of the market and are attributed to the same trader category. Similar to Veredas et al. (2002), we find that the amount of contemporaneous observations is around 20-25% of the original sample, depending on the stock. Clearly, our approach does not solve the arbitrariness issue related to null durations, though at least it takes into account the three critical attributes of a trade: time, market direction, and trader identity (category). Starting from trade durations, we complement our analysis with two additional types of durations that have drawn the attention of microstructure researchers:

- *Price durations* are defined as the time necessary to observe a cumulative price change not less than a certain threshold. More precisely, a price duration is generated by selecting two points in time such that:

  $$|p_i - p_{i-k}| \geq c$$

  where p_i represents the transaction price, c is an exogenous constant used as reference threshold, and $k \geq 1$. As it was stressed by Engle and Russell (1998), the study of price durations is particularly appealing because of their relationship with instantaneous volatility. We adopt the midquote in place of transaction price to avoid the bias caused by the bid-ask bounce (Roll, 1984). We set the values of the

[3]The data set provides a variable to determine the side of the market which initiates a transaction.

threshold c equal to 0.015 and 0.025 Euro, which correspond to three and five times the minimum tick size (0.005 Euro) allowed at Euronext Paris. We label the first case 'Low threshold' and the second one 'High threshold'.

- *Volume durations* are defined as the time necessary to trade a specific aggregate volume. Volume durations are generated by retaining two following points in time, $i-k$ and i, such that:

$$\sum_{i-k}^{i} volume_i \geq c$$

  where c represents the exogenous threshold for the amount of traded shares. The interest for volume durations is justified as they represent a simple proxy for market liquidity: a long volume duration is indicative of a market characterized by a low level of liquidity, while the opposite holds for short volume durations. For each stock, we choose a threshold c which is equal to 10 and 20 times the average number of exchanged shares. As before, we label the first case 'Low threshold' and the second one 'High threshold'.

4.4.2 Descriptive statistics

Some descriptive statistics for trade, price, and volume durations are summarized in Table 4.1. For trade durations, the table evidences a coefficient of variation (CV) larger than the one reported in other researches, though some stocks in Hautsch (2004) present similar values. Such a CV confirms the overdispersion that is typical of trade durations, as well as the inadequacy of the exponential distribution to fit these data. A plausible explanation for the high CV may be related to the minimum value of trade durations, which is apparently small for all the stocks. Indeed, the use of microseconds assures a higher level of precision, but it increases the proportion of extremely small durations and the overall dispersion of the series. Contrary to previous empirical researches, volume durations exhibit a vague underdispersion, which is actually apparent only for the high threshold case. Indeed, the same stylized fact is not observed for low-threshold volume durations[4]. Volume durations present a remarkably high autocorrelation, sometimes displaying an autocorrelation coefficient ρ larger than 0.5 at the first lag. This is not unexpected and it may be partially attributed to volume clustering itself; in any case, the values are in line with other contributions (e.g. Hautsch, 2004, and Bauwens and Veredas, 2004). As to price durations, they prove to be considerably overdispersed, as in Engle and Russell (1998) or Bauwens and Giot (2003). Moreover, price durations also exhibit a considerable autocorrelation pattern, though the ρ coefficient is generally lower, especially for the high threshold case. Given the remarkable value of the first autocorrelation coefficients, it is not surprising to find a strong rejection of the null of serial uncorrelation for all types of durations at the first 30 lags.

The descriptive analysis is supplemented by Figures 4.1 and 4.2 that provide an illustrative comparison of the financial durations examined in this paper. We choose Alstom as representative stock, but similar patterns are also exhibited by the remaining series. The upper panel of Figure 4.1 displays the time series plots which emphasize the clustering behaviour common to all kinds of durations. Figure 4.1 also confirms the parallelism between (G)ARCH and ACD models, and it evidences the generally high level of data dispersion. The lower panel of Figure 4.1 shows the 30-lag correlogram that highlights the exceptionally slow decline of the autocorrelation function. In some cases, several lags are necessary before attaining a non-significant autocorrelation value and Figure 4.1 makes clear the different persistence among the three durations. In fact, in spite of a higher starting value, the correlogram of volume durations tends toward zero at a faster rate with respect to the one of price and trade durations.

In the upper panel of Figure 4.2, we provide the empirical density estimation, computed through Epanechnicov kernel with optimal bandwidth. To make the figure intelligible, we reduce the scale to a maximum value of 1200 seconds, even though the upper

[4] In the preliminary analysis of data, we have tried different thresholds for volume durations, and we have noticed that the degree of underdispersion increases by augmenting the value of c. This matches with the descriptive statistics provided in Hautsch (2004) for Eurex and ASX, where volume durations turn out to be underdispersed when the aggregation threshold is raised.

Table 4.1: Descriptive statistics for trade, price and volume durations. For each type of duration, the table reports the number of observations, the average, the standard deviation, the coefficient of variation (CV), the minimum, the maximum, and the 1-lag and 10-lag autocorrelation coefficient.

	Trade durations				
	Alstom	Axa	Crédit A.	Sanofi-A.	Schneider
Observations	160,170	230,073	187,947	247,581	153,729
Avg. Duration	8.43	5.87	7.18	5.46	8.78
St. Dev.	16.99	10.91	13.99	10.47	18.89
CV	2.02	1.86	1.95	1.92	2.15
Min	0.002	0.002	0.002	0.002	0.002
Max	498.96	293.46	315.75	280.04	379.05
ρ_1	0.20	0.20	0.19	0.19	0.17
ρ_{10}	0.10	0.12	0.10	0.10	0.09

	Price durations - Low threshold				
	Alstom	Axa	Crédit A.	Sanofi-A.	Schneider
Observations	33,664	13,771	11,855	30,464	35,173
Avg. Duration	40.01	96.70	111.47	44.27	38.34
St. Dev.	62.53	154.14	173.65	69.50	60.52
CV	1.56	1.59	1.56	1.57	1.58
Min	0.012	0.044	0.018	0.012	0.012
Max	1218.90	2636.40	2527.50	2314.00	1915.30
ρ_1	0.21	0.29	0.26	0.23	0.20
ρ_{10}	0.15	0.18	0.15	0.14	0.13

	Price durations - High threshold				
	Alstom	Axa	Crédit A.	Sanofi-A.	Schneider
Observations	22,142	6,680	11,855	18,032	26,870
Avg. Duration	60.62	198.33	111.47	74.82	49.99
St. Dev.	92.74	308.46	173.65	113.57	78.65
CV	1.53	1.56	1.56	1.52	1.57
Min	0.012	0.096	0.018	0.016	0.012
Max	1605.60	6308.50	2527.50	2885.40	2191.60
ρ_1	0.23	0.29	0.26	0.27	0.21
ρ_{10}	0.16	0.15	0.15	0.14	0.15

	Volume durations - Low threshold				
	Alstom	Axa	Crédit A.	Sanofi-A.	Schneider
Observations	13,985	19,815	16,474	21,361	13,790
Avg. Duration	104.08	73.47	88.87	68.22	106.27
St. Dev.	116.03	79.81	94.55	72.05	116.81
CV	1.11	1.09	1.06	1.06	1.09
Min	0.032	0.024	0.028	0.060	0.052
Max	1186.30	1119.20	1284.70	862.16	1351.80
ρ_1	0.49	0.50	0.47	0.49	0.47
ρ_{10}	0.24	0.26	0.25	0.28	0.23

Table 4.1: Continued from the previous page.

	Volume durations - High threshold				
	Alstom	Axa	Crédit A.	Sanofi-A.	Schneider
Observations	7,412	10,579	8,775	11,570	7,065
Avg. Duration	189.60	132.81	160.27	121.50	198.35
St. Dev.	189.70	130.40	151.43	114.59	192.48
CV	1.00	0.98	0.94	0.94	0.97
Min	0.064	0.068	0.084	0.058	0.176
Max	1834.00	1643.50	1649.90	1081.90	1886.90
ρ_1	0.52	0.55	0.50	0.55	0.52
ρ_{10}	0.22	0.27	0.24	0.30	0.21

bound of some series is definitely higher. All the durations exhibit a non-surprising right-skewed shape, which is representative of the high concentration of relatively small durations. This feature is particularly evident for trade durations, while it is less marked for price and volume durations because of the aggregation effect. The lower panel of Figure 4.2 displays the intraday seasonality plot that is estimated by means of a cubic spline function. The presence of some intraday cycles is a standard commonality of several microstructure variables, such as spread, volume, and volatility, and it has been widely discussed in previous researches. Figure 4.2 shows the typical inverted U-shape pattern, with a more intense trading activity in the early morning and in the late afternoon, and a drop of the trading frequency around midday. The intraday periodicity directly affects the degree of autocorrelation of the series, and several approaches have been proposed to take this effect into account.

As an example, Andersen and Bollerslev (1998) suggest to model the seasonal trend through a Fourier series approximation, while Veredas et al. (2002) advance a semiparametric method which makes use of the Nadaraya-Watson kernel to jointly estimate the seasonal component and the parameters of the ACD model. In this paper, we follow the original, and probably most popular, approach by Engle and Russell (1998) who assume

Figure 4.1: Time series plot and correlogram for trade, price and volume durations. The figure displays the high threshold series for aggregate durations; all graphs refer to Alstom.

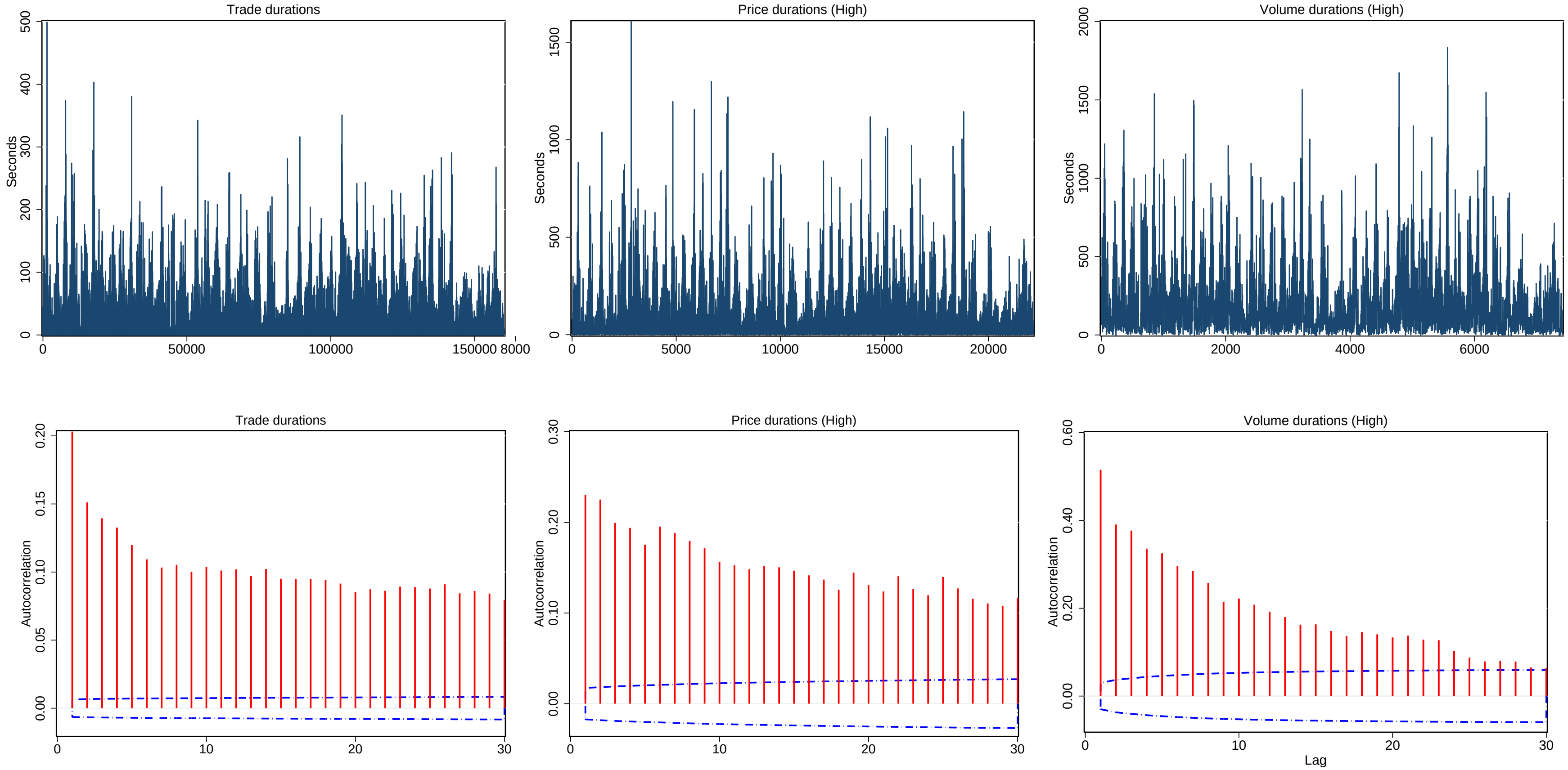
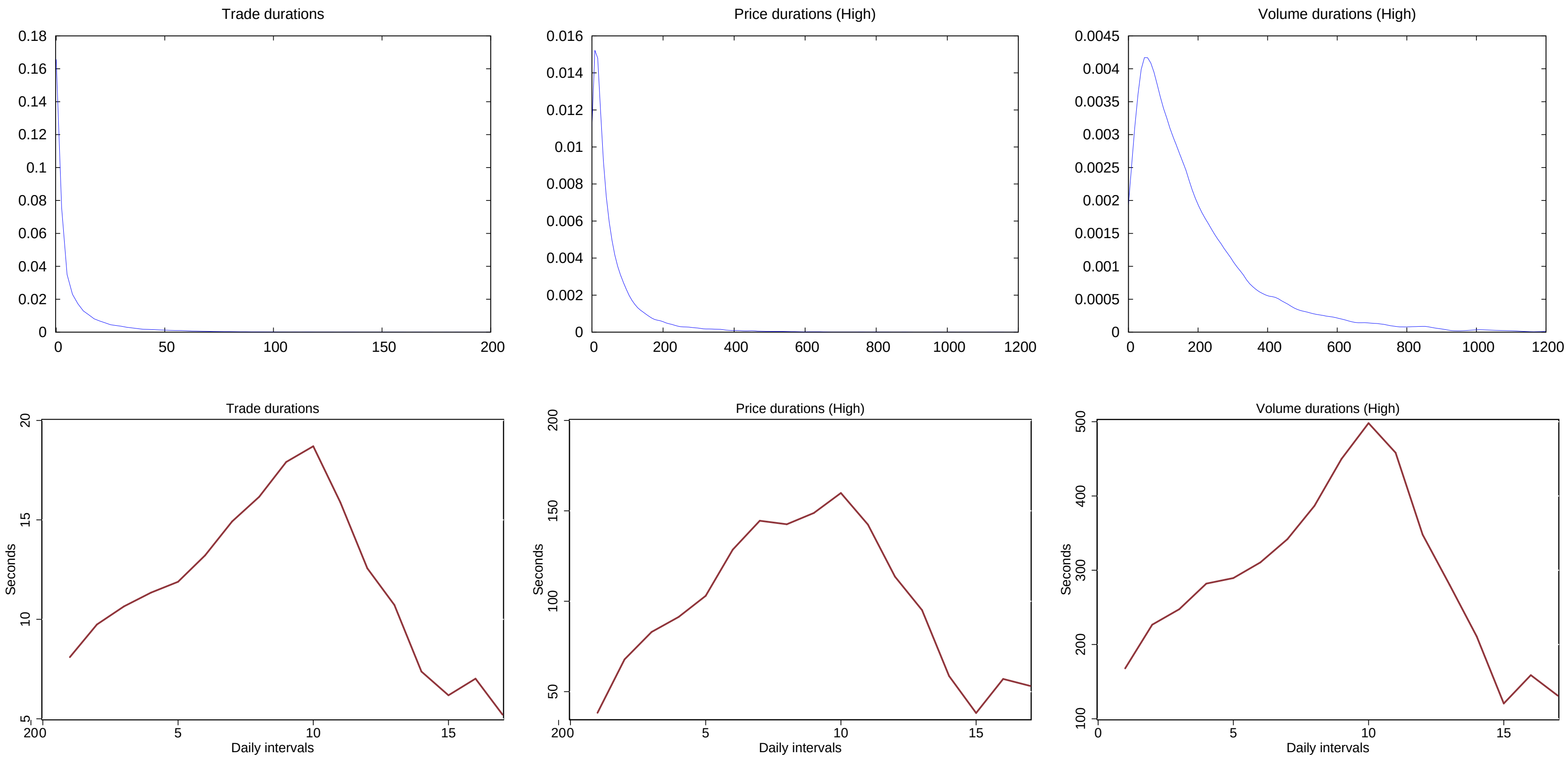


Figure 4.2: Density estimate and seasonality plot for trade, price and volume durations. The figure displays the high threshold series for aggregate durations; all graphs refer to Alstom.



a deterministic seasonal component affecting the duration series multiplicatively:

$$d_i = x_i s(i)$$

where d_i is a generic financial duration, x_i is the corresponding seasonally adjusted variable, and $s(i)$ is the seasonality component which is function of time i. Also in this case, we refer to x_i as the seasonally adjusted variable, independent of a specific type of duration. To proceed in the estimation, we first compute the expectation of d_i conditional on the time of day, i.e. we average d_i over 30-minute intervals; then, we use a cubic spline to filter out the seasonal effect[5]. We choose a two-step procedure and we separately filter out the series before estimating ML parameters. When the number of observations is sufficiently large, Engle and Russell (1998) show that this approach does not significantly affect the results. All the results presented in Section 5 and 6 are based on the deseasonalized series x_i.

4.5 Empirical specification

In this section, we describe our strategy to augment the specification of ψ in the Log-ACD model, in order to study the impact of traders on financial durations. As was previously discussed, the Log-ACD model does not require any positivity constraint on the conditional mean parameters, which makes it particularly suited for the testing of microstructure hypotheses. We choose a distinct specification for trade durations with respect to price and volume durations; this comes as a result of using aggregate series too.

[5] We apply this method to extract the intraday cycle from all the variables displaying such a similar pattern, like the spread or the volume.

4.5.1 Trade durations

In the case of trade durations, we add some explanatory variables to ψ_i along the scheme presented in Tables 4.2 and 4.3. The former shows the transaction volume and the information on trader identities expressed as a categorical variable; the latter displays the straightforward combination of the two variables. As an example, InfoB equals the transaction volume when the buyer is an informed agent, i.e. he exhibits a '2' or a '7' as agent code; on the other hand, LiqprovS equals the transaction volume when the seller coincides with the liquidity provider (code '6'). A similar association also holds for the remaining cases. All these variables are meant to measure the volume-weighted impact on time of the different categories of investors. We believe that the combination of transaction volume and traders' identities represents a more reliable approach than simply considering the categorical dummies for market participants; this choice is motivated by the acknowledged informativeness of the trading volume (Easley and O'Hara, 1987). As to the lag structure, we adopt a Log-ACD$_1$ model with two lags for x_i and one lag for ψ_i. Preliminary estimates have found in this specification the optimal compromise among computation time, information criteria, model parsimony, and autocorrelation issues; however, our results are qualitatively similar even with the more parsimonious Log-ACD$_1$ (1,1) model. In the end, the new specification of ψ_i for trade durations becomes:

$$\begin{aligned} \psi_i = \omega + \alpha_1 \ln x_{i-1} + \alpha_2 \ln x_{i-2} + (\beta - \alpha_1 - \alpha_2)\psi_{i-1} \\ +\delta_1\, InfoB_{i-1} + \delta_2\, LiqprovB_{i-1} + \delta_3\, InfoS_{i-1} + \delta_4\, LiqprovS_{i-1} \end{aligned} \tag{4.7}$$

where δ_1, δ_2, δ_3, and δ_4 are the parameters that measure the impact of traders on time. Clearly, we exclude from Equation (4.9) the variables measuring the contribution of uninformed traders to avoid collinearity. Independently of the market side, we expect both informed traders and the liquidity provider to reduce the time intervals between consecutive transactions, though we believe that this effect is driven by distinct motivations. For the liquidity provider, the speeding up of the trading process may be easily attributed to

Table 4.2: Transaction volumes and dummies for traders' identities. Data are from Alstom.

Volume	Buyer			Seller		
	Uninf.	Inf.	Liq. Provider	Uninf.	Inf.	Liq. Provider
530	0	1	0	1	0	0
1,000	1	0	0	0	1	0
15	0	0	1	1	0	0
78	0	1	0	0	0	1

Table 4.3: Transaction volumes and trader combinations. Data are from Alstom.

Volume	Buyer			Seller		
	UninfoB	InfoB	LiqprovB	UninfoS	InfoS	LiqprovS
530	0	530	0	530	0	0
1,000	1,000	0	0	0	1,000	0
15	0	0	15	15	0	0
78	0	78	0	0	0	78

his contractual duties. On the other hand, a similar impact for informed traders is more likely explained by means of theoretical information models such as Easley and O'Hara (1987) or Easley and O'Hara (1992). When entering the market, informed traders try to quickly maximize their informational privilege via a trading strategy which combines an intense market activity with a high transaction volume.

4.5.2 The case of aggregate durations

Unfortunately, the previous approach is not suitable to be pursued for price and volume durations, at least for two reasons. First of all, we generally collapse several observations to define the aggregate series, which is fairly easy to appreciate from the descriptive statistics displayed in Table 4.1. In this case, we cannot assess the impact of traders on each single transaction by following the same method proposed for trade durations. We solve this issue by computing, for each spell, the fraction of volume attributed to each category of investor. As an example, we measure the contribution of informed traders on the buy side as:

$$InfoB_i = \frac{Buy\,Informed Volume_i}{Total\,volume_i}$$

where, with a little abuse of notation, we denote with $Buy\,InformedVolume_i$ the total transaction volume exchanged by informed buyers during the aggregated duration x_i, while $Total\,volume_i$ represents the overall number of shares exchanged during the same interval. Similar definitions are straightforward to derive for the remaining categories of traders, as well as for the sell side of the market. Generally speaking, we measure the impact on time by means of a volume-weighted proxy of the traders' activity.

The second reason to modify the specification of Equation (4.9) still concerns the aggregation criterion employed to define price and volume durations. Actually, a dissimilar number of transactions is normally required to generate each spell of aggregate durations; this is clearly evident from Figure 4.2 which shows the seasonal pattern during the continuous auction. Neglecting this point is a failing strategy, as it disregards the informational content of transaction frequency within each spell. As an example, a volume duration generated through a single 'informed' transaction is clearly different from a volume duration which collapses several small-sized transactions executed by distinct categories of traders. This is in line with Easley and O'Hara (1992) who theoretically recognize the relationship between the presence of informed traders in the market and the intensity of trading. Therefore, we adapt the structure of Equation (4.7) to characterize the state of the market on the basis of the trading frequency. To identify two alternative regimes, we define the dummy variable $\xi_i = 1$ when the number of transactions in the spell x_i is larger than the average of the whole series; this almost splits the distribution of the number of transactions per spell into two fairly equal partitions[6]. Our strategy partially follows Wong et al. (2008), though their paper investigates a completely different hypothesis.

[6]The frequency distribution of price durations is generally left-skewed. We estimate the Log-ACD models for price durations by also applying the median as a threshold for ξ_i, and we obtain qualitatively similar results. Therefore, we keep a homogeneous approach in the whole sample and adopt the average number of transactions for price durations too.

Then, we make use of ξ_i to propose the following nonlinear specification of ψ_i:

$$\begin{aligned} \psi_i = \omega + \alpha_1 \ln x_{i-1} + \alpha_2 \ln x_{i-2} + (\beta - \alpha_1 - \alpha_2)\psi_{i-1} + \\ \xi_{i-1}(\delta_1\, InfoBH_{i-1} + \delta_2\, LiqprovBH_{i-1} + \delta_3\, InfoSH_{i-1} + \delta_4\, LiqprovSH_{i-1}) + \\ (1 - \xi_{i-1})(\delta_5\, InfoBL_{i-1} + \delta_6\, LiqprovBL_{i-1} + \delta_7\, InfoSL_{i-1} + \delta_8\, LiqprovSL_{i-1}) \end{aligned} \tag{4.8}$$

where we add the suffix $-H$ and $-L$ to distinguish between the two states of the market. In this way, we duplicate the specification of Equation (4.7) in order to test the existence of a regime-specific impact of traders[7]. For informed traders, we expect a negative impact on aggregate durations when the market is found in an intense state of trading activity. In fact, on the basis of the theoretical models, the information-driven trading should be effective only when it is associated with a high frequency of transactions. On the other hand, and because of his contractual duties, the liquidity provider should significantly reduce the length of a spell under both market regimes. As a consequence of our specification, we should therefore expect a negative impact when the average trading size is lower. This is quite obvious for volume durations given that the threshold c adopted to define a series is fixed; however, simple descriptive statistics generally confirm the inverse relationship between trading intensity and transaction size for price durations too. At a first sight, this appears to contradict the hypothesis on volume informativeness (see Easley and O'Hara, 1987), but this is not actually the case. In fact, the informational content of volume has to be distinctly evaluated *within* each state of the market, not *between*. The variables measuring the weighted participation of traders are successful in doing that, as they ponder the presence of traders within each of the two market regimes.

[7]Coherent with the specification adopted for trade durations, we exclude the contribution of uninformed traders from Equation (4.8) to avoid collinearity in the estimation.

4.6 Results

This section analyses the ML estimates of the basic specifications defined in Section 4.5. The structure outlined in Equation (4.7) represents the reference model for trade durations, while Equation (4.8) provides the empirical specification for price and volume durations. To save space, we only report the variables that measure the impact of traders on financial durations; the complete list of results is available upon request. The estimates are based on the Burr and the Generalized Gamma distributions[8], and all our considerations rely on a significance level which is adjusted according to the sample size. For trade durations, we use an α-level equal to 1%, while we adopt a significance threshold equal to 5% for the smaller samples of aggregate durations. In this way, we minimize the possibility that a large sample size radically affects the significance of our results, see Hausman et al. (1992). On the whole, this choice is coherent with previous empirical researches, and sometimes it appears even more restrictive (see Hautsch, 2003 or Bauwens and Veredas, 2004). Table 4.4 displays the results for trade durations and with respect to both innovation densities. The estimates for price and volume durations based on the Burr distribution are collected in Table 4.5, while the results based on the Generalized Gamma distribution are provided in Table 4.6.

4.6.1 Trade durations

In the case of trade durations, the estimates exhibit a striking uniformity across the whole sample and with respect to both distributions taken into account for ML estimation. Table 4.4 shows that an increase in the volume-weighted activity generated by informed traders and the liquidity provider reduces the inter-trade spell and accelerates the trading frequency. This result holds for both sides of the market and it may be easily appreciated by looking at the negative coefficients of $InfoB$, $InfoS$, $LiqprovB$, and

[8]We omit the results based on the exponential density, though they are generally in line with the ones derived through the more flexible distributions. Our choice is motivated by the strong misspecification displayed by the exponential density when it is tested along the lines discussed in Section 4.7. This is not totally unexpected given the high level of overdispersion exhibited by the data.

Table 4.4: ML estimates for trade durations, with robust standard errors in parentheses.

Trade durations - Burr distribution						
	Alstom		Axa		Crédit A.	
	Coeff.	(Std. Error)	Coeff.	(Std. Error)	Coeff.	(Std. Error)
InfoB	-1.13E-02***	(1.28E-03)	-1.26E-03***	(2.72E-04)	-3.50E-03***	(3.43E-04)
LiqprovB	-1.95E-02***	(1.85E-03)	-6.06E-03***	(4.16E-04)	-6.89E-03***	(4.47E-04)
InfoS	-1.96E-02***	(1.26E-03)	-3.16E-03***	(2.73E-04)	-4.16E-03***	(3.40E-04)
LiqprovS	-2.45E-02***	(1.67E-03)	-8.94E-03***	(4.44E-04)	-7.35E-03***	(5.37E-04)

Trade durations - Burr distribution				
	Sanofi-A.		Schneider	
	Coeff.	(Std. Error)	Coeff.	(Std. Error)
InfoB	-7.82E-03***	(1.23E-03)	-6.41E-05***	(2.07E-05)
LiqprovB	-2.29E-02***	(2.27E-03)	-1.41E-04***	(4.34E-05)
InfoS	-1.38E-02***	(1.39E-03)	-1.65E-04***	(2.30E-05)
LiqprovS	-3.04E-02***	(2.47E-03)	-3.14E-04***	(4.22E-05)

Trade durations - Generalized Gamma distribution						
	Alstom		Axa		Crédit A.	
	Coeff.	(Std. Error)	Coeff.	(Std. Error)	Coeff.	(Std. Error)
InfoB	-1.22E-02***	(1.40E-03)	-2.16E-03***	(4.04E-04)	-3.96E-03***	(6.22E-04)
LiqprovB	-2.06E-02***	(2.01E-03)	-7.79E-03***	(8.53E-04)	-8.06E-03***	(1.06E-03)
InfoS	-2.13E-02***	(1.39E-03)	-4.19E-03***	(4.20E-04)	-4.63E-03***	(5.97E-04)
LiqprovS	-2.56E-02***	(1.84E-03)	-1.04E-02***	(6.81E-04)	-8.07E-03***	(1.06E-03)

Trade durations - Generalized Gamma distribution				
	Sanofi-A.		Schneider	
	Coeff.	(Std. Error)	Coeff.	(Std. Error)
InfoB	-8.83E-03***	(1.56E-03)	-6.14E-05***	(2.16E-05)
LiqprovB	-2.77E-02***	(4.94E-03)	-1.56E-04***	(4.53E-05)
InfoS	-1.58E-02***	(2.27E-03)	-1.60E-04***	(2.44E-05)
LiqprovS	-3.47E-02***	(4.62E-03)	-3.17E-04***	(4.71E-05)

LiqprovS. Coherent with the discussion in Section 4.5.1, we believe that two distinct explanations exist for these findings. On the one hand, the liquidity provider enters the LOB to facilitate the transactions among market operators and our estimates confirm his effective role in the provision of market liquidity. On the other hand, the clustering pattern generated by informed-based trading is more likely explained by the theoretical information models as Easley and O'Hara (1992). In this latter case, two motivations concur to the negative impact associated with $InfoB$ and $InfoS$. First, informed agents hurry to exploit their short-lived private signal; thus, they trade more aggressively in terms of submission frequency. Second, a large informed trade is generally matched with multiple orders to be fully executed; the consequence of this splitting is a sequence of small and almost simultaneous transactions. Clearly, some overlap subsists between these two motivations, but we leave the analysis of the prevailing effect behind informed-based trading for future research.

4.6.2 Aggregate durations

The nonlinear specification defined in Equation (4.8) provides evidence of a regime-specific effect for aggregate durations. According to Tables 4.5 and 4.6, when the market is found in the fast-trading regime, a more intense activity coming from informed traders and the liquidity provider has a negative impact on time. On the contrary, such effect is generally not observed for the state of the market characterized by a low trading frequency. This result is a bit more evident for the sell side of the market and it may be appreciated by looking at the negative and significant coefficients of $InfoBH$, *LiqprovBH*, $InfoSH$, and *LiqprovSH*. On the whole, these findings comply with the estimates presented for trade durations[9].

[9] Although Easley and O'Hara (1992) directly refer to transaction frequency, their model also extends to price and volume durations which actually represent an aggregation of multiple transactions (see Bauwens and Giot, 2003 or Bauwens and Veredas, 2004).

4.6.2.1 Price durations

As to price durations, *LiqprovSH* is the only variable that always displays a negative and statistically significant coefficient. On the other hand, and considering both price series, $InfoBH$ is negative and significant in the 70% of cases, *LiqprovBH* in the 60% (50%) of cases for Generalized Gamma (Burr) distribution, and $InfoSH$ in the 80% of cases. No coefficients exhibit a positive and significant sign when the market is found in the intense trading regime. Conversely, the large majority of the estimates are not statistically significant when the market is characterized by a low frequency of transactions. In a few marginal cases, we find a positive and statistically significant coefficient for $InfoBL$ and a negative one for *LiqprovSL*; nevertheless, these estimates are generally quite close to the critical value, and an unambiguous pattern may be spotted at most for *LiqprovSL*. On the whole, Tables 4.5 and 4.6 confirm the existence of a trader-related effect for price durations too, though this time, the impact turns out to be highly dependent on the intensity of the market activity. These findings may be worth an additional interpretation on the basis of the inverse relationship between price durations and instantaneous volatility (Engle and Russell, 1998). The motivation for a higher volatility in the market is generally twofold. On the one hand, the high volatility can be explained through the noise produced by the liquidity traders, on the other hand it can be attributed to a great concentration of informed traders in the market (Dufour and Engle, 2000a, Wong et al., 2008). According to the sign of $InfoBH$ and $InfoSH$, the estimates for price durations seem to validate the second hypothesis. In other words, a higher presence of informed traders reduces the length of price durations, and it increases the instantaneous volatility because of an informational-driven pressure. On the contrary, we believe that the estimates for the liquidity provider are more easily associated with his role in ensuring market liquidity, especially at less favorable price levels. This explanation seems more plausible when looking at the negative and significant coefficients that are sometimes encountered for *LiqprovSL*.

Table 4.5: ML estimates for price and volume durations; robust standard errors in parentheses. The estimates are based on the Burr distribution.

Price durations - Low threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.059**	(2.68E-02)	-0.068	(3.91E-02)	-0.100**	(4.87E-02)	-0.058***	(1.97E-02)	-0.058***	(1.28E-02)
LiqprovBH	-0.050	(3.25E-02)	-0.097	(5.76E-02)	-0.014	(6.42E-02)	-0.077***	(3.09E-02)	-0.079***	(2.19E-02)
InfoSH	-0.089***	(2.63E-02)	-0.116***	(2.68E-02)	-0.112***	(4.78E-02)	-0.063**	(2.04E-02)	-0.032**	(1.32E-02)
LiqprovSH	-0.145***	(3.78E-02)	-0.109***	(3.81E-02)	-0.191***	(6.67E-02)	-0.104***	(3.37E-02)	-0.065***	(2.25E-02)
InfoBL	-0.006	(2.91E-02)	0.061**	(3.03E-02)	0.049	(2.70E-02)	0.013	(1.11E-02)	0.023**	(9.56E-03)
LiqprovBL	0.077	(4.02E-02)	0.034	(3.07E-02)	0.004	(2.77E-02)	0.021	(1.34E-02)	0.023	(1.30E-02)
InfoSL	-0.037	(3.09E-02)	-0.006	(4.85E-02)	0.021	(2.68E-02)	-0.012	(1.09E-02)	0.004	(9.47E-03)
LiqprovSL	-0.079	(4.95E-02)	-0.059	(5.18E-02)	-0.016	(2.69E-02)	-0.034**	(1.41E-02)	-0.026**	(1.31E-02)

Price durations - High threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.060***	(2.06E-02)	-0.069	(8.61E-02)	-0.047	(7.90E-02)	-0.074**	(3.41E-02)	-0.077***	(1.98E-02)
LiqprovBH	-0.062	(3.71E-02)	-0.314**	(1.16E-01)	-0.008	(1.02E-01)	-0.112**	(5.29E-02)	-0.077**	(3.64E-02)
InfoSH	-0.076***	(2.24E-02)	-0.345***	(8.86E-02)	-0.166**	(7.81E-02)	-0.061	(3.60E-02)	-0.019	(2.04E-02)
LiqprovSH	-0.129***	(3.40E-02)	-0.283**	(1.25E-01)	-0.426***	(1.03E-01)	-0.184***	(5.61E-02)	-0.081***	(3.77E-02)
InfoBL	0.005	(1.22E-02)	0.002	(4.81E-02)	-0.031	(5.36E-02)	0.024	(2.02E-02)	0.025**	(1.05E-02)
LiqprovBL	0.014	(9.78E-03)	-0.026	(5.34E-02)	-0.003	(5.81E-02)	0.058**	(2.47E-02)	0.014	(1.40E-02)
InfoSL	0.003	(1.70E-02)	-0.065	(5.04E-02)	0.038	(5.46E-02)	-0.034	(1.92E-02)	-0.003	(1.04E-02)
LiqprovSL	-0.032	(2.18E-02)	-0.085	(5.60E-02)	0.089	(5.85E-02)	-0.022	(2.55E-02)	-0.029**	(1.44E-02)

Volume durations - Low threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.046***	(1.63E-02)	-0.015	(1.58E-02)	-0.107***	(1.64E-02)	-0.041***	(1.31E-02)	-0.052***	(1.95E-02)
LiqprovBH	-0.051**	(2.43E-02)	-0.017	(2.13E-02)	-0.043**	(2.04E-02)	-0.048**	(1.55E-02)	-0.041	(3.49E-02)
InfoSH	-0.050***	(1.56E-02)	-0.038**	(1.59E-02)	-0.014	(4.71E-02)	-0.033***	(1.30E-02)	-0.013	(2.03E-02)
LiqprovSH	-0.122***	(2.37E-02)	-0.105***	(2.17E-02)	-0.030	(5.29E-02)	-0.063***	(2.09E-02)	-0.124***	(3.70E-02)
InfoBL	0.020	(1.46E-02)	0.012	(8.48E-03)	0.023**	(9.28E-03)	0.006	(9.76E-03)	0.026	(1.38E-02)
LiqprovBL	0.042	(2.47E-02)	-0.009	(1.26E-02)	0.007	(1.64E-02)	-0.024	(1.32E-02)	0.005	(2.19E-02)
InfoSL	-0.017	(1.44E-02)	0.018**	(8.11E-03)	0.017	(1.15E-02)	0.015	(1.14E-02)	-0.010	(1.35E-02)
LiqprovSL	0.008	(2.37E-02)	0.016	(1.38E-02)	-0.068***	(1.48E-02)	-0.069	(1.86E-02)	-0.065***	(2.49E-02)

Volume durations - High threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.055**	(2.81E-02)	-0.024	(1.44E-02)	-7.95E-02***	(2.94E-02)	-0.036**	(1.49E-02)	-0.041	(3.34E-02)
LiqprovBH	-0.050	(3.91E-02)	0.007	(3.00E-02)	-6.49E-02	(4.38E-02)	-0.089***	(3.06E-02)	-0.071	(6.48E-02)
InfoSH	-0.091***	(3.01E-02)	-0.069***	(1.83E-02)	-2.11E-02	(2.16E-02)	-0.041**	(1.84E-02)	-0.032	(3.33E-02)
LiqprovSH	-0.149***	(4.00E-02)	-0.141***	(2.90E-02)	-7.68E-02**	(3.84E-02)	-0.033	(3.00E-02)	-0.190***	(6.65E-02)
InfoBL	-0.008	(2.96E-02)	0.017	(1.24E-02)	2.77E-02	(2.36E-02)	-0.009	(2.20E-02)	0.014	(2.47E-02)
LiqprovBL	0.075	(4.33E-02)	0.004	(2.95E-02)	3.29E-02	(3.94E-02)	0.030	(2.11E-02)	-0.039	(3.72E-02)
InfoSL	-0.036	(2.92E-02)	0.016	(1.13E-02)	3.53E-02	(2.52E-02)	0.020	(1.51E-02)	-0.021	(2.44E-02)
LiqprovSL	-0.073	(4.55E-02)	0.046	(2.95E-02)	-1.10E-01***	(3.61E-02)	-0.103***	(3.82E-02)	-0.075	(4.43E-02)

Table 4.6: ML estimates for price and volume durations; robust standard errors in parentheses. The estimates are based on the Generalized Gamma distribution.

Price durations - Low threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.049***	(1.79E-02)	-0.065	(3.56E-02)	-0.106**	(5.21E-02)	-0.053**	(2.01E-02)	-0.058**	(1.30E-02)
LiqprovBH	-0.037	(2.72E-02)	-0.101**	(4.46E-02)	-0.042	(7.28E-02)	-0.074**	(3.12E-02)	-0.077**	(2.23E-02)
InfoSH	-0.078***	(1.74E-02)	-0.120***	(3.27E-02)	-0.113***	(5.23E-02)	-0.062***	(2.08E-02)	-0.032***	(1.34E-02)
LiqprovSH	-0.108***	(2.65E-02)	-0.110**	(4.69E-02)	-0.190***	(7.27E-02)	-0.099***	(3.41E-02)	-0.060***	(2.30E-02)
InfoBL	0.010	(8.98E-03)	0.064**	(2.43E-02)	0.051	(2.63E-02)	0.016	(1.12E-02)	0.023**	(9.73E-03)
LiqprovBL	0.013	(1.01E-02)	0.038	(2.52E-02)	0.008	(2.71E-02)	0.021	(1.36E-02)	0.024	(1.33E-02)
InfoSL	0.003	(8.66E-03)	-0.009	(2.89E-02)	0.002	(2.59E-02)	-0.010	(1.11E-02)	0.003	(9.66E-03)
LiqprovSL	-0.039***	(9.76E-03)	-0.059	(3.11E-02)	-0.025	(2.63E-02)	-0.032**	(1.43E-02)	-0.024	(1.34E-02)

Price durations - High threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.059**	(2.98E-02)	-0.074	(8.73E-02)	-0.047	(7.83E-02)	-0.071**	(3.49E-02)	-0.076***	(2.01E-02)
LiqprovBH	-0.061	(7.71E-02)	-0.349***	(1.15E-01)	-0.009	(9.95E-02)	-0.108**	(5.33E-02)	-0.076**	(3.70E-02)
InfoSH	-0.080***	(2.51E-02)	-0.362***	(9.02E-02)	-0.153**	(7.61E-02)	-0.056	(3.66E-02)	-0.021	(2.08E-02)
LiqprovSH	-0.132**	(5.58E-02)	-0.248**	(1.25E-01)	-0.425***	(1.03E-01)	-0.174***	(5.67E-02)	-0.077**	(3.85E-02)
InfoBL	0.003	(1.46E-02)	0.010	(4.91E-02)	-0.028	(5.15E-02)	0.033	(2.04E-02)	0.024**	(1.07E-02)
LiqprovBL	0.016	(2.71E-02)	-0.031	(5.46E-02)	-0.009	(5.74E-02)	0.063**	(2.48E-02)	0.015	(1.43E-02)
InfoSL	0.004	(2.21E-02)	-0.086	(5.15E-02)	0.040	(5.36E-02)	-0.033	(1.93E-02)	-0.002	(1.05E-02)
LiqprovSL	-0.029	(4.70E-02)	-0.091	(5.71E-02)	0.083	(5.82E-02)	-0.018	(2.56E-02)	-0.028	(1.47E-02)

Volume durations - Low threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.050**	(2.20E-02)	-0.013	(1.54E-02)	-0.116***	(2.67E-02)	-0.044**	(1.38E-02)	-0.060***	(1.93E-02)
LiqprovBH	-0.060**	(2.75E-02)	-0.019	(2.11E-02)	-0.049***	(1.68E-02)	-0.049**	(1.73E-02)	-0.054	(3.49E-02)
InfoSH	-0.053***	(1.58E-02)	-0.037**	(1.53E-02)	-0.010	(2.52E-02)	-0.035***	(1.41E-02)	-0.015	(2.02E-02)
LiqprovSH	-0.140***	(2.63E-02)	-0.110***	(2.12E-02)	-0.037	(2.69E-02)	-0.074***	(2.63E-02)	-0.141***	(3.70E-02)
InfoBL	0.009	(2.12E-02)	0.019**	(7.79E-03)	0.022	(1.35E-02)	0.009	(1.13E-02)	0.024	(1.26E-02)
LiqprovBL	0.034	(4.79E-02)	-0.010	(1.20E-02)	0.006	(1.58E-02)	-0.036	(2.20E-02)	0.020	(1.93E-02)
InfoSL	-0.029	(2.16E-02)	0.015**	(7.09E-03)	0.012	(6.42E-03)	0.012	(7.97E-03)	-0.018	(1.20E-02)
LiqprovSL	0.014	(3.53E-02)	0.009	(1.29E-02)	-0.076***	(1.46E-02)	-0.075	(2.55E-02)	-0.074***	(2.22E-02)

Volume durations - High threshold										
	Alstom		Axa		Crédit A.		Sanofi-A.		Schneider	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
InfoBH	-0.070**	(2.80E-02)	-0.026	(3.27E-02)	-8.45E-02***	(3.02E-02)	-0.037	(2.04E-02)	-0.032	(3.09E-02)
LiqprovBH	-0.032	(3.77E-02)	0.018	(1.83E-02)	-6.95E-02**	(3.48E-02)	-0.087***	(3.04E-02)	-0.104	(6.23E-02)
InfoSH	-0.093***	(2.54E-02)	-0.070***	(2.49E-02)	-1.99E-02	(2.17E-02)	-0.057**	(2.44E-02)	-0.041	(3.16E-02)
LiqprovSH	-0.170***	(3.81E-02)	-0.146***	(3.26E-02)	-8.81E-02***	(3.31E-02)	-0.033	(2.91E-02)	-0.196***	(6.50E-02)
InfoBL	0.011	(1.66E-02)	0.019	(2.06E-02)	3.24E-02	(1.83E-02)	0.002	(2.28E-02)	0.002	(2.22E-02)
LiqprovBL	0.049	(4.10E-02)	0.016	(1.95E-02)	3.75E-02	(2.84E-02)	0.009	(3.19E-02)	0.009	(3.01E-02)
InfoSL	-0.055	(3.25E-02)	0.021	(1.92E-02)	3.25E-02	(1.92E-02)	0.011	(1.45E-02)	-0.025	(2.13E-02)
LiqprovSL	-0.088	(4.73E-02)	0.039	(2.76E-02)	-1.15E-01***	(3.97E-02)	-0.117***	(4.02E-02)	-0.105***	(4.01E-02)

4.6.2.2 Volume durations

By examining the ML estimates for volume durations, we notice that $InfoBH$ is negative and significant in 60% (70%) of cases for Generalized Gamma (Burr) distribution, $LiqprovBH$ in the 50% (40%) for Generalized Gamma (Burr) distribution, $InfoSH$ in the 60%, and $LiqprovSH$ in the 80%. Contrary to price durations, the effect of $LiqprovSL$ seems more robust, as this variable displays a significant coefficient in 50% (40%) of the estimates based on the Generalized Gamma (Burr) distribution. No unambiguous effect can be identified for the few remaining variables exhibiting a significant coefficient. To interpret the results, we recall that volume durations may be thought of as a proxy of market liquidity. According to Tables 4.5 and 4.6, the liquidity provider is once more acknowledged as an accelerator of the exchanges in the market, sometimes even during the periods of slow trading activity. As before, we believe that this result is more likely attributed to his contractual constraints. The explanation for informed traders is clearly different, and we emphasize that our estimates are significant when the average transaction volume is lower. In our opinion, this result comes as a direct consequence of volume informativeness. When full information is not disclosed in the market, an unexpected large order could be revealing of a private signal. Informed traders prefer to disguise their identity by submitting close sequences of smaller size orders, which partially explains the strong clustering pattern displayed by volume durations. This kind of trading strategies is analyzed in the stealth trading literature (see Chakravarty, 2001) which finds that informed traders submit medium size orders to hide their private information. Hautsch (2004) adapts the stealth trading hypothesis to the study of trade durations, but he does not find any convincing result. In our empirical specification, we do not distinguish among classes of trade size as in Chakravarty (2001), and we check the robustness of our findings along the lines discussed in the next section. However, we recognize a certain empirical soundness, and we leave the relationship between financial durations and stealth trading literature as a vein for future research.

4.7 Robustness checks and diagnostics

In this section, we examine some robustness checks to minimize the possibility that our findings are mainly driven by the huge size of the data set. We consider some augmented specifications for all the series, and we especially focus on aggregated durations as they are traditionally devoted to the testing of microstructure hypotheses (e.g. Bauwens and Giot, 2000 or Bauwens and Giot, 2003).

4.7.1 Trade durations

We start with trade durations and we augment Equation (4.7) as:

$$\begin{aligned} \psi_i = \omega + \alpha_1 \ln x_{i-1} + \alpha_2 \ln x_{i-2} + (\beta - \alpha_1 - \alpha_2)\psi_{i-1} \\ +\delta_1\, InfoB_{i-1} + \delta_2\, LiqprovB_{i-1} + \delta_3\, InfoS_{i-1} + \delta_4\, LiqprovS_{i-1} \\ +\delta_5\, Dspread_{i-1} + \delta_6\, Big_{i-1} + \delta_7\, InitS_{i-1} \end{aligned} \tag{4.9}$$

where $Dspread_{i-1} = spread_{i-1} - spread_{i-2}$, Big is a dummy variable indicating a transaction size larger than the average, and $InitS$ is a dummy variable that identifies the transactions initiated by the sell side of the market. We use *Dspread* to measure the impact of the bid-ask spread on financial durations. Some theoretical contributions (e.g. Glosten and Milgrom, 1985 or Easley and O'Hara, 1992) have illustrated the relationship between the width of the spread and the presence of informed traders in the market. Informed traders foster the trading frequency in order to exploit their private signal; in this context, the market maker widens the spread to protect himself from the risk of a loss due to informed-based trading. Therefore, we expect a negative relationship between time and spread, though this hypothesis has been mainly tested in quote-driven markets and for aggregate durations. Big is used to spot particularly large transactions, in order to check the microstructure hypothesis that assumes a negative relationship between trade durations and transaction volume (Easley and O'Hara, 1987). This is because a large transaction volume is generally perceived as a signal for the presence of informed traders in the market; coherent with the previous point, a higher percentage of informed traders is associated with an increasing frequency of transactions. Finally, $InitS$ is employed to check for the existence of an asymmetric effect between the two sides of the market. As an example, if there exists a panic effect related to clustered sequences of sell transactions, then we expect a negative coefficient for this variable.

Table 4.7 provides the estimates of the robust specification presented in Equation (4.9). To save space, we limit the exposition to the results obtained for Alstom; however, the rest of the sample displays similar findings and the complete set of estimates is available upon request. First of all, we notice that the previous results for trade durations are fully validated, both in terms of significance and of direction of the effect. Secondly, as to the additional variables, we find a partial confirmation of the previous microstructure hypotheses. The coefficient for $Dspread$ is always positive and significant for all the stocks, which is contradictory to theoretical models on informed trading. Nevertheless, our results are in line with Hautsch (2004) who also find a positive coefficient for the bid-ask spread. As an explanation for that, he suggests that the widening of the spread discourages the market activity because of higher transaction costs. On the other hand, we find a negative and significant estimate for Big, which is in line with the assumption that a larger volume accelerates the trading frequency (Wong et al., 2008 or Hautsch, 2004). Finally, we generally do not find a significant estimate for $InitS$, which excludes the presence of an asymmetric effect between the two sides of the market[10].

4.7.2 Aggregate durations

As to aggregated durations, the selection of additional regressors is facilitated by the previous empirical research that widely concentrates on the testing of microstructure hypotheses, see Bauwens and Giot (2003) or Bauwens and Veredas (2004). In this paper,

[10] We also try a specification where we disentangle the transaction volume from the trader identity, to exclude that our findings are mainly driven by the negative impact attributed to the trade size. The results are in line with Section 4.6.1. To save space, we do not report the estimates, also because we consider this specification as less relevant, provided that it does not "weight" the activity of traders with the volume informational content.

Table 4.7: ML estimates for trade durations; robust standard errors in parentheses. The results refer to the augmented specification defined in Section 4.7. The table provides the estimates for Alstom.

	Trade durations			
	Burr distribution		G.G. distribution	
	Coeff	Std.Error	Coeff	Std.Error
InfoB	-4.69E-05***	(9.73E-06)	-4.60E-05***	(1.00E-05)
LiqprovB	-6.62E-05***	(1.48E-05)	-6.07E-05***	(1.53E-05)
InfoS	-7.56E-05***	(9.45E-06)	-7.38E-05***	(9.73E-06)
LiqprovS	-1.05E-04***	(1.44E-05)	-9.42E-05***	(1.49E-05)
DSpread	3.01E-01***	(3.07E-03)	3.17E-01***	(3.03E-03)
Big	-2.12E-02***	(4.11E-03)	-2.16E-02***	(4.23E-03)
InitS	-1.81E-03	(2.97E-03)	-1.79E-03	(3.05E-03)

we limit the analysis to two supplementary variables and for price durations we augment the specification introduced in Equation (4.8) as follows:

$$\begin{aligned} \psi_i = \omega + \alpha_1 \ln x_{i-1} + \alpha_2 \ln x_{i-2} + (\beta - \alpha_1 - \alpha_2)\psi_{i-1} + \\ \xi_{i-1}(\delta_1\, InfoBH_{i-1} + \delta_2\, LiqprovBH_{i-1} + \delta_3\, InfoSH_{i-1} + \delta_4\, LiqprovSH_{i-1}) + \\ (1-\xi_{i-1})(\delta_5\, InfoBL_{i-1} + \delta_6\, LiqprovBL_{i-1} + \delta_7\, InfoSL_{i-1} + \delta_8\, LiqprovSL_{i-1}) + \\ +\delta_9\, Dspread_{i-1} + \delta_{10}\, Avgvol_{i-1} \end{aligned} \tag{4.10}$$

where $Avgvol_{i-1}$ represents the average transaction volume in the previous spell and $Dspread_{i-1}$ is defined as before. Contrary to other empirical contributions (Bauwens and Giot, 2000, Bauwens and Giot, 2003), we exclude the number of transactions occurring in the previous spell from the variables to be tested; nevertheless, a simpler version of this regressor is implicitly taken into account through the nonlinear specification adopted for ψ_i. The inverse relationship between transaction volume and price durations is justified by the theoretical information models discussed in Section 4.7.1, and it has been confirmed by plenty of empirical contributions (e.g. Bauwens and Veredas, 2004). Coherent with the previous literature, we expect a higher average volume to speed up the price adjustment process.

For volume durations, we distinguish the contribution of the supplementary regressors on the basis of the trading activity. Indeed, some preliminary results show that even the additional microstructure variables may be affected by the state of the market. Therefore, we augment the basic specification as follows:

$$
\begin{aligned}
\psi_i = \omega + \alpha_1 \ln x_{i-1} + \alpha_2 \ln x_{i-2} + (\beta - \alpha_1 - \alpha_2)\psi_{i-1} + \\
\xi_{i-1}(\delta_1\, InfoBH_{i-1} + \delta_2\, LiqprovBH_{i-1} + \delta_3\, InfoSH_{i-1} + \delta_4\, LiqprovSH_{i-1} + \\
+\delta_9\, DspreadH_{i-1} + \delta_{10}\, DmidpointH_{i-1}) + \\
(1-\xi_{i-1})(\delta_5\, InfoBL_{i-1} + \delta_6\, LiqprovBL_{i-1} + \delta_7\, InfoSL_{i-1} + \delta_8\, LiqprovSL_{i-1} + \\
+\delta_{11}\, DspreadL_{i-1} + \delta_{12}\, DmidpointL_{i-1})
\end{aligned} \tag{4.11}
$$

where $DmidpointH_{i-1} = midpoint_{i-1} - midpoint_{i-2}$ represents the absolute midquote change, and, as before, we use the suffix $-H$ and $-L$ to differentiate the two states of the market. *Dmidpoint* measures the price impact over each volume duration, and it is associated with the market reaction curve . Given a specific volume size, the price impact is an increasing function of the amount of information in the market, and we expect a negative effect of *Dmidpoint* on the length of volume durations. In fact, when the price impact is higher, the market should be characterized by an intense activity coming from informed traders, see Hautsch (2003).

The estimates based on the robust specifications for price and volume durations are reported in Tables 4.8 and 4.9. We do not spend too much effort in a detailed analysis of the two tables, and we limit our discussion to a few principal results. First of all, also in this case, we recognize that our findings are in line with the estimates displayed in Table 4.5 and 4.6 for the basic model. Second, and particularly interesting, we observe that the additional regressors present a different outcome for price and volume durations. As to price durations, Tables 4.8 and 4.9 actually display the expected negative sign for *Dspread* and for *Avgvol*. The sign of *Avgvol* is in line with the estimates of trade durations, while the impact of spread turns out to be negative, similar to Engle and Russell (1998) or Bauwens and Veredas (2004); in this case, the informational effect related to the bid-ask spread seems to prevail over the increase of transaction costs. Conversely, the effect of *Dspread* and *Dmidpoint* is mostly found to be non-significant for volume durations. Given the surprising result, we further investigate on this point and we re-estimate the

Table 4.8: ML estimates for price and volume durations; robust standard errors in parentheses. The results are based on the Burr distribution and they refer to the augmented specifications defined in Section 4.7. The tables provide the estimates for Alstom.

	Price durations			
	Low threshold		High threshold	
	Coeff	Std.Error	Coeff	Std.Error
InfoBH	-4.90E-02***	(1.65E-02)	-6.21E-02**	(2.60E-02)
LiqprovBH	-4.21E-02	(2.39E-02)	-6.39E-02	(3.99E-02)
InfoSH	-8.08E-02***	(1.64E-02)	-7.61E-02***	(2.54E-02)
LiqprovSH	-1.20E-01***	(2.36E-02)	-1.39E-01***	(3.78E-02)
InfoBL	1.51E-02	(1.02E-02)	3.47E-03	(1.26E-02)
LiqprovBL	1.96E-02	(1.15E-02)	1.71E-02	(1.43E-02)
InfoSL	8.54E-03	(9.84E-03)	2.60E-03	(1.22E-02)
LiqprovSL	-3.79E-02***	(1.12E-02)	-3.19E-02**	(1.39E-02)
Dspread	-3.70E-02***	(2.94E-03)	-3.67E-02***	(3.40E-03)
Avgvol	-8.16E-05***	(2.06E-05)	-5.88E-05**	(2.77E-05)

	Volume durations			
	Low threshold		High threshold	
	Coeff	Std.Error	Coeff	Std.Error
InfoBH	-3.94E-02**	(1.84E-02)	-5.02E-02**	(2.90E-02)
LiqprovBH	-4.95E-02**	(2.44E-02)	-5.25E-02	(5.52E-02)
InfoSH	-3.81E-02**	(1.41E-02)	-8.79E-02***	(3.05E-02)
LiqprovSH	-1.20E-01***	(2.36E-02)	-1.53E-01***	(4.70E-02)
DspreadH	1.30E-02	(2.00E-02)	-1.41E-02	(3.30E-02)
DmidpointH	3.13E-02	(3.76E-02)	6.45E-02	(4.36E-02)
InfoBL	-2.97E-02	(3.07E-02)	-4.20E-02	(2.75E-02)
LiqprovBL	-8.83E-03	(3.41E-02)	-8.66E-02	(4.55E-02)
InfoSL	-6.03E-01	(4.56E-01)	-6.21E-01	(5.22E-01)
LiqprovSL	-1.94E-02	(3.01E-02)	1.16E-01	(1.65E-01)
DspreadL	2.64E-01	(4.46E-01)	-3.31E-01	(5.07E-01)
DmidpointL	2.21E-01	(1.56E-01)	2.90E-01	(1.57E-01)

	Volume durations - Microstructure variables			
	Low threshold		High threshold	
	Coeff	Std.Error	Coeff	Std.Error
DspreadH	-8.36E-01***	(2.61E-01)	-9.69E-01***	(3.24E-01)
DmidpointH	-3.10E-01***	(9.26E-02)	-1.86E-01**	(8.69E-02)
DspreadL	1.44E+00***	(2.77E-01)	7.55E-01**	(3.68E-01)
DmidpointL	3.85E-01***	(1.24E-01)	4.64E-01***	(1.39E-01)

Table 4.9: ML estimates for price and volume durations; robust standard errors in parentheses. The results are based on the Generalized Gamma distribution and they refer to the augmented specifications defined in Section 4.7. The tables provide the estimates for Alstom.

	Price durations			
	Low threshold		High threshold	
	Coeff	Std.Error	Coeff	Std.Error
InfoBH	-4.95E-02***	(1.67E-02)	-6.02E-02**	(2.67E-02)
LiqprovBH	-3.96E-02	(2.43E-02)	-6.32E-02	(4.08E-02)
InfoSH	-7.98E-02***	(1.65E-02)	-8.00E-02***	(2.60E-02)
LiqprovSH	-1.13E-01***	(2.40E-02)	-1.42E-01***	(3.84E-02)
InfoBL	1.52E-02	(1.03E-02)	1.25E-03	(1.28E-02)
LiqprovBL	2.12E-02	(1.17E-02)	1.85E-02	(1.46E-02)
InfoSL	8.42E-03	(9.98E-03)	3.46E-03	(1.25E-02)
LiqprovSL	-3.56E-02***	(1.13E-02)	-2.80E-02**	(1.42E-02)
Dspread	-3.67E-02***	(2.98E-03)	-3.62E-02***	(3.45E-03)
Avgvol	-7.79E-05***	(2.10E-05)	-5.85E-05**	(2.82E-05)

	Volume durations			
	Low threshold		High threshold	
	Coeff	Std.Error	Coeff	Std.Error
InfoBH	-4.35E-02**	(1.82E-02)	-6.78E-02**	(3.30E-02)
LiqprovBH	-5.83E-02**	(2.60E-02)	-3.87E-02	(5.18E-02)
InfoSH	-4.24E-02***	(1.47E-02)	-9.33E-02***	(3.26E-02)
LiqprovSH	-1.40E-01***	(2.28E-02)	-1.79E-01***	(4.01E-02)
DspreadH	1.15E-03	(2.09E-02)	7.39E-03	(3.58E-02)
DmidpointH	2.04E-02	(2.22E-02)	3.84E-02	(4.68E-02)
InfoBL	-3.89E-02	(2.23E-02)	-5.88E-02	(3.68E-02)
LiqprovBL	-1.01E-03	(3.42E-02)	-1.03E-01	(6.35E-02)
InfoSL	-7.21E-01**	(3.18E-01)	-6.87E-01	(5.67E-01)
LiqprovSL	-4.91E-02	(1.27E-01)	1.13E-01	(1.98E-01)
DspreadL	1.18E-01	(3.57E-01)	-6.72E-01	(6.90E-01)
DmidpointL	1.79E-01	(1.72E-01)	2.64E-01	(1.63E-01)

	Volume durations-Microstructure variables			
	Low threshold		High threshold	
	Coeff	Std.Error	Coeff	Std.Error
DspreadH	-8.84E-01**	(3.36E-01)	-1.04E+00***	(3.59E-01)
DmidpointH	-3.66E-01***	(1.42E-01)	-2.17E-01**	(1.03E-01)
DspreadL	1.33E+00***	(3.78E-01)	5.33E-01	(2.79E-01)
DmidpointL	3.70E-01**	(1.80E-01)	4.49E-01***	(1.24E-01)

model by only taking into account *Dspread* and *Dmidpoint* in the specification of ψ; the results are reported in the last panel of Tables 4.8 and 4.9. The two tables show that *Dspread* and *Dmidpoint* are significant only when we exclude the variables for trader impact from the definition of the conditional duration.

This is particularly interesting, as it attributes a dominant explicative power to the fraction of volume traded by informed traders and the liquidity provider. Furthermore, the coefficients at the bottom of Tables 4.8 and 4.9 display an opposite result according to the state of the market. Indeed, the increase of *Dspread* and *Dmidpoint* negatively impacts the future volume spell only when the trading intensity is high. On the contrary, the same variables reduce or marginally influence the demand for market liquidity in the slow trading regime. In other words, a wider spread or an increasing price impact affect the length of volume durations only when the informed-based trading is more likely to occur; otherwise, both variables lose most of their significance. On the whole, these findings are coherent, though unexpected, with the theoretical models that consider the relationship between the trading frequency and the presence of informed agents in the market (Easley and O'Hara, 1992).

4.7.3 Residuals

We end this section with a brief examination of residual diagnostics based on the density forecast technique. We follow the same approach introduced by Diebold et al. (1998) in the context of GARCH models, and we refer to Bauwens et al. (2004) for a more exhaustive coverage of the topic. Diebold et al. (1998) essentially propose to test the correct specification of a model by means of the probability integral transforms:

$$z_i = \int_{-\infty}^{x_i} f_i(u)du \tag{4.12}$$

where $f_i(u)$ represents the density forecast implied by the duration model. Under the assumption of correct specification, the distribution of the empirical sequence of probability integral transform is IID $U(0,1)$. We show the results of residual diagnostics through a set of z-histograms, as it is common practice in the previous empirical contributions. To save space, we limit the analysis to Alstom, but the rest of the sample exhibit a similar outcome.

A few comments are in order with respect to the six plots presented in Figure 4.3. First of all, we notice a quite remarkable similitude between the z-histograms based on the Burr and on the Generalized Gamma distributions, such that a superior performance of one of the two distributions is not clearly spot. Second, the z-histograms for trade durations display a slightly pronounced U-shaped pattern, putting too much weight on the tails of the distribution and over-representing the very small and the very high durations. The z-histograms for aggregated durations appear more uniformly distributed, with the only exception of the very high durations which are generally over-represented. On the whole, our z-histograms are in line with the plots showed in some previous works, as Bauwens et al. (2004) or Vuorenmaa (2011); nevertheless, a formal χ_2 test for goodness of fit accepts the two distributions only in a few marginal cases.

Finally, we use the sequence of probability integral transforms to check for the absence of serial correlation in model residuals, and we report the graphs for Alstom in Figure 4.4. With the only exception of trade durations, the plots show that our models manage to extract most of the serial dependence from the original series, which is also confirmed by the Ljung-Box statistics for the first 30 lags. We believe that the opposite findings for trade and aggregate durations should be attributed to the high level of persistence displayed by trade durations, which is moreover emphasized by the use of microseconds. Indeed, the presence of microseconds assures a more accurate measurement of time, but it also increases the autocorrelation of the series.

Figure 4.3: Z-histograms for trade, volume, and price durations. The figure displays the high threshold series for aggregate durations; all graphs refer to Alstom.

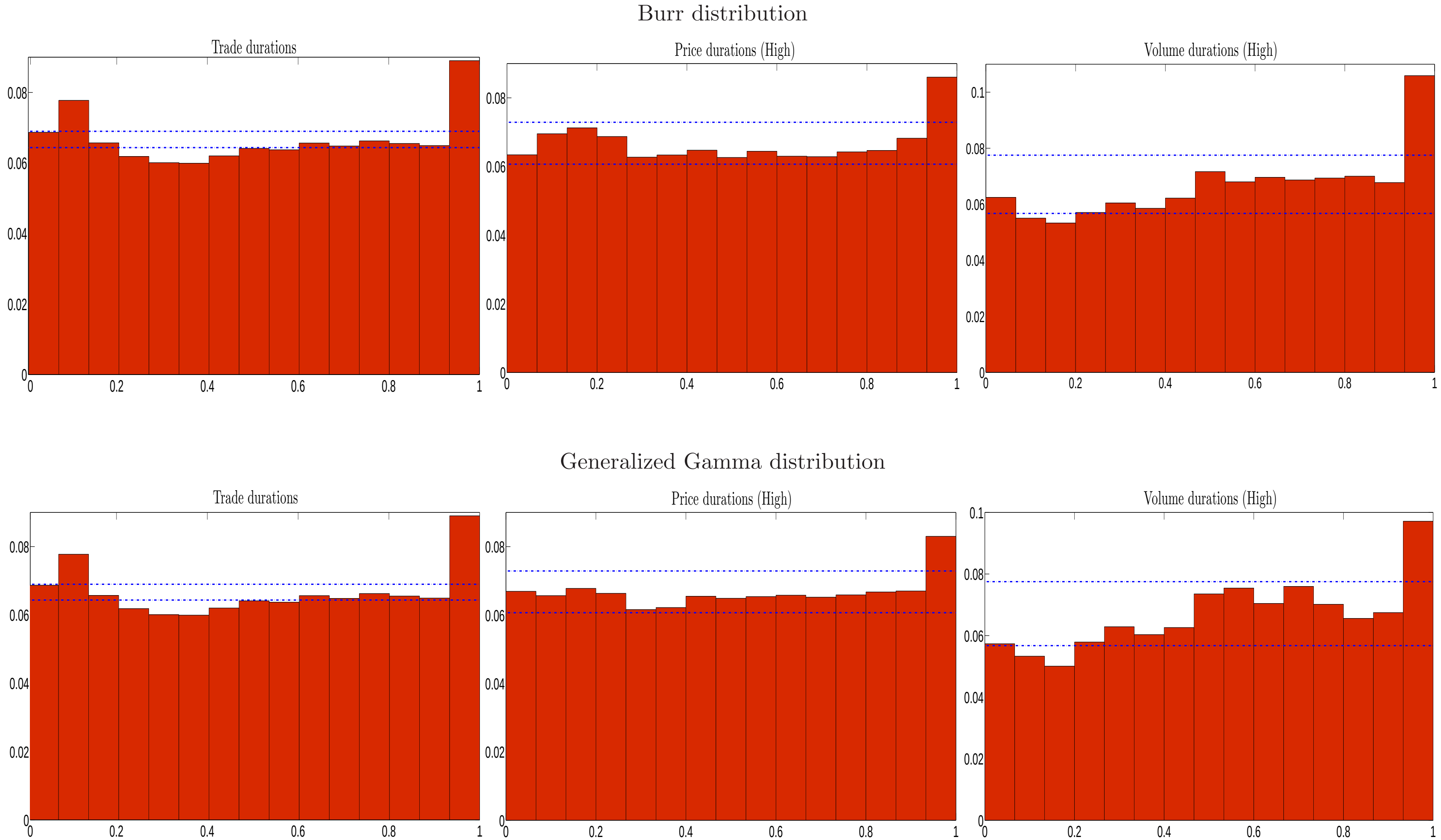
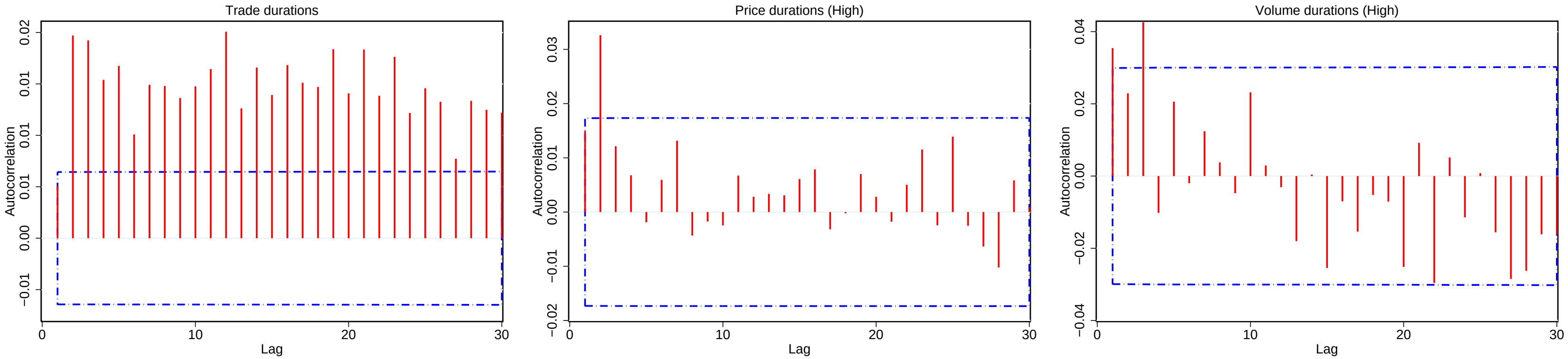
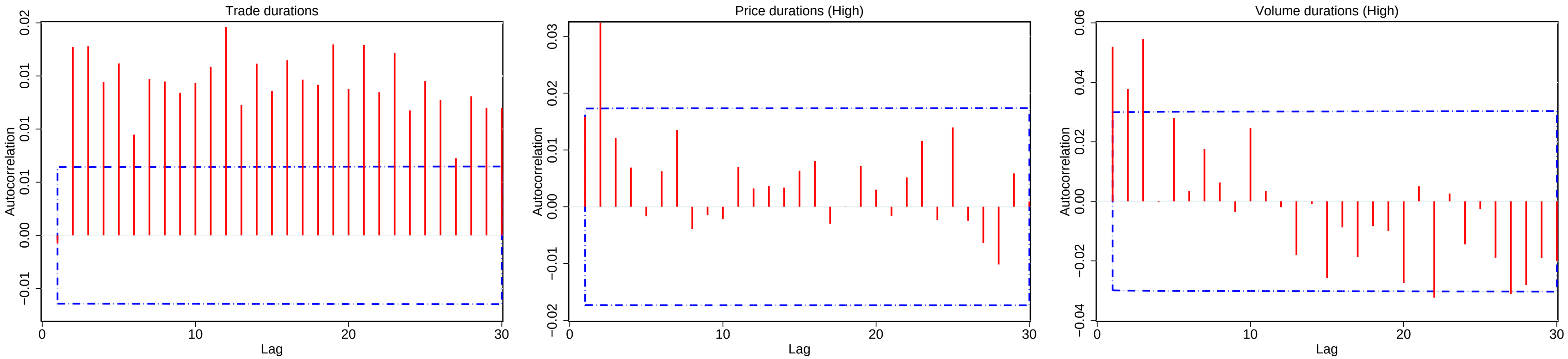


Figure 4.4: Z-residual correlogram for trade, price and volume durations. The figure displays the high threshold series for aggregate durations; all graphs refer to Alstom.

Burr distribution



Generalized Gamma distribution



4.8 Conclusions

In this research, we adopt a Log-ACD model to analyse the impact of traders on financial durations. We test our hypothesis on trade, price, and volume durations, and we find that informed traders and the liquidity provider foster the arrival of future market events. This result is unequivocally true for trade durations, while for price and volume durations, our estimates are proven to depend on the state of the market activity. Indeed, the impact of informed traders and the liquidity provider is effective when their presence is combined with a high trading frequency. On the whole, our estimates are coherent with the standard theoretical models discussed in financial market microstructure, and they confirm the dominant role of informed traders and of the liquidity provider in driving the market activity. The results are robust across alternative distributions, as well as when they are tested with supplementary microstructure variables. On the other hand, residual diagnostics show that the lag specification and the distributional assumptions are only partially successful in fitting the duration series.

Bibliography

Allen, D., Chan, F., McAleer, M., and Peiris, S. (2008), "Finite Sample Properties of the QMLE for the Log-ACD model: Application to Australian Stocks," *Journal of Econometrics*, 147, 1, 163-183.

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

Bauwens, L. (2006), "Econometric Analysis of Intra-Daily Trading Activity on the Tokyo Stock Exchange," *Monetary and Economic Studies*, 24 (1), 1-24.

Bauwens, L., Galli, F., and Giot, P. (2003), "The Moments of Log-ACD Model," *CORE Discussion Paper*, No. 2003/11.

Bauwens, L., and Giot, P. (2000), "The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process of Three NYSE Stocks," *Annales d'Economie et de Statistique*, 60, 117-149.

Bauwens, L., and Giot, P. (2001), "Econometric Modelling of Stock Market Intraday Activity," *Kluwer Academic Publishing.*

Bauwens, L., and Giot, P. (2003), "Asymmetric ACD Models: Introducing Price Information in ACD Models," *Empirical Economics*, 28, 709-731.

Bauwens, L., Giot, P., Grammig, J., and Veredas, D. (2004), "A Comparison of Financial

Duration Models via Density Forecasts," *International Journal of forecasting*, 20 (4), 589-609.

Bauwens, L. and D. Veredas (2004), "The Stochastic Conditional Duration Model: a Latent Variable Model for the Analysis of Financial Durations," *Journal of Econometrics*, 119 (2), 381-482.

Biais, B., Hillion, P., and Spatt, C. (1995), "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *The Journal of Finance*, 50, 1655-1689.

Chakravarty, S. (2001), "Stealth Trading: which Traders' Trades Move Stock Prices?," *Journal of Financial Economics*, 61, 289-307.

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

Diebold, F.X., Gunther, T.A., and Tay, A.S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.

Dufour, A., and Engle, R.F. (2000a), "The ACD Model: Predictability of the Time between Consecutive Trades," Discussion Paper, ISMA Centre, University of Reading.

Dufour, A., and Engle, R.F. (2000b), "Time and the Impact of a Trade" *Journal of Finance* 55, 2467-2498.

Easley, D., and O'Hara, M. (1987), "Price, Trade and Information in Securities Market," *Journal of Financial Economics*, 19, 69-90.

Easley, D., and O'Hara, M. (1992), "Time and the Process of Security Price Adjustment," *The Journal of Finance*, 47, 577-605.

Engle, R.F. (1982), "Autoregressive Conditional Heteroschedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

Engle, R.F., (2000), "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68, 1-22.

Engle, R.F., and Russell, J.R. (1998), "Autoregressive Conditional Duration: a new Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127-1162.

Ferriani, F. (2012), "Informed and Uninformed Traders at Work: Evidence from the French Market," *Munich Repec working paper*.

Foucault, T., Moinas, S., and Theissen, E. (2007), " Does Anonymity Matter in Electronic Limit Order Markets?," *The Review of Financial Studies*, 5, 1707-1747.

Glosten, L., and Milgrom, J. (1985), "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71-100.

Gourieroux, C., Jasiak, J. and Le Fol, G. (1999), "Intra-Day Market Activity," *Journal of Financial Markets*, 2, 193-226.

Grammig, J., and Maurer, K.O. (2000), "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model," *Econometrics Journal*, 3, 16-38.

Hasbrouck, J. (1991), "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, 179-209.

Hausman, J., Lo, A. and MacKinlay, A.C. (1992), "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319-379.

Hautsch, N. (2003), "Assessing the Risk of Liquidity Suppliers on the Basis of Excess Demand Intensities," *Journal of Financial Econometrics*, 1, 189-215.

Hautsch, N. (2004), "Modelling Irregularly Spaced Financial Data-Theory and Practice of Dynamic Duration Models," *Lecture Notes in Economics and Mathematical Systems*, Springer Berlin.

Lee, C.M.C., and Ready, M.J. (1991), "Inferring Trade Direction from Intraday Data," *The Journal of Finance*, 46, 733-746.

Lunde, A. (1999), "A Generalized Gamma Autoregressive Conditional Duration Model," *discussion paper* Aalborg University.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 349-370.

Pacurar, M. (2008), "Autoregressive Conditional Duration Models in Finance: a Survey of the Theoretical and Empirical Literature," *Journal of Economic Surveys*, 22, 4, 711-751.

Roll, R. (1984), "A Simple Implicit Measure of the Bid-Ask Spread in an Efficient Market," *The Journal of Finance*, 39, 1127-39.

Veredas, D., Rodriguez-Poo, J., and Espasa, A., (2002), "On the (Intradaily) Seasonality of a Financial Point Process," *CORE DP 2002/23*, Université catholique de Louvain, Louvain-La-Neuve.

Vuorenmaa, T.A., (2011), "Q-Weibull Autoregressive Conditional Duration Model and Threshold Dependence," *working paper*.

Wong, W.K., Tan, D., and Tian, Y., (2008), "Informed Trading and Liquidity in the Shanghai Stock Exchange," *International Review of Financial Analysis*, 18, 66-73.

Zhang, M.Y., Russell, J.R., and Tsay, R.S. (2001), "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data," *Journal of Econometrics*, 104, 179-207.

Chapter 5

Far Away from the Best: Order Aggressiveness at Euronext Paris

Fabrizio Ferriani

JEL codes: C10, G10, G14

Keywords: Market microstructure, Order aggressiveness, Limit order book, Simultaneous equation model, Euronext Paris

5.1 Introduction

In an order-driven market, a trader essentially submits two typologies of orders that generate an opposite impact on the liquidity of the limit order book (LOB). A trader enhances the market liquidity by choosing a limit order, and conversely subtracts liquidity from the LOB by submitting a market order. The diffusion of order-driven markets worldwide has increased the attention of market microstructure research on the determinants of order submission strategies. Generally speaking, traders decide which order to place on the basis of a trade-off between profitability and the probability of execution. When a market order is submitted to the LOB, then it is immediately executed; however, the matching may occur at unfavourable execution prices, especially in illiquid markets.

On the other hand, the traders satisfy their price preferences in the case of limit orders, but they lose the certainty of execution; in fact, there could be no proposals in the LOB to meet their price constraints. Limit orders also face an adverse selection risk, given that their prices are fixed and market monitoring is a costly activity. As a matter of fact, when new information arrives in the market, the informed traders may profit from their private signal by picking off a standing limit order that is then executed at unfavourable price conditions (Copeland and Galai, 1983, Biais et al., 1995).

Biais et al. (1995), hereafter BHS (1995), analyse the Paris Stock Exchange, and the relationship between the order flow and the state of the LOB. They propose a ranking where market and limit orders are classified according to different levels of aggressiveness. Order aggressiveness represents a measure of traders' impatience and it is directly related to the probability of execution. The most aggressive traders submit market orders or marketable limit orders that are promptly matched and guarantee an immediate execution. Conversely, the less impatient traders place limit orders that are associated with progressively decreasing levels of aggressiveness: as an extreme case, the lowest level of aggressiveness collects buy (ask) orders with a limit price below (above) the best bid (ask) quote. BHS's (1995) seminal contribution has widely influenced the following research on empirical market microstructure, and their scheme has been used to test order aggressiveness in several financial markets, especially by means of discrete response models. As a few examples, their approach is adopted by Ranaldo (2004) in the analysis of the Swiss Stock Exchange, while Ellul et al. (2003) and Cao et al. (2009) replicate BHS's (1995) classification for the NYSE and the Australian Stock Exchange, respectively. All these papers examine order aggressiveness through ordered probit or multinomial logit models, which are easily implemented and also allow to deal with price discreteness issues.

An in-depth investigation of order aggressiveness is also provided by Duong et al. (2009) who extend the previous research by taking into account the categories of traders. They use an ordered probit model to examine the order flow at the Australian Stock Exchange, differentiating between institutional and individual (retail) order flows. Pascual and Veredas (2009) propose a two-stage sequential ordered probit to study order aggressiveness at the Spanish Stock Exchange. In the first step, each trader is classified with respect to the decision of providing or consuming market liquidity. In the second step, the trader chooses the best submission strategy according to the level of aggressiveness determined previously. Hall and Hautsch (2006) adopt BHS's (1995) ranking to examine the order arrival process through an autoregressive conditional intensity model, and they find support for a multivariate dynamics of order arrivals; in a way, their paper follows the previous research by Bisière and Kamionka (2000). Lo and Sapp (2010) present a simultaneous equation model to evaluate order aggressiveness in the currency markets. Their paper is sensibly appealing for two reasons. First, they consider a market that is traditionally neglected by the empirical microstructure literature. Second, and also more relevant, they split order aggressiveness into two components: price and quantity. The price component is clearly dominant in determining the traders' impatience given that markets are traditionally organized in order to enhance a strict price priority. However, all conditions being equal, previous theoretical contributions (e.g. Easley and O'Hara, 1987) recognize a strong informational content to the trading volume. On this point, Bae et al. (2003) stress the relevance of quantity for the analysis of order submission strategies, and they suggest the examination of its relationship with market microstructure variables.

In this research, we partially replicate Lo and Sapp (2010) approach to study order aggressiveness, and we investigate the impact of LOB variables on the price-quantity decision at Euronext Paris. This paper is not the first attempt to study order aggressiveness at Euronext Paris, since BHS (1995) seminal contribution or Bisière and Kamionka (2000) also focus on the French Stock Exchange. Nevertheless, our data set presents some peculiarities that are worth a more in-depth investigation. A first point concerns the market reform introduced on 23 April 2001, when Euronext Paris switched to an anonymous trading regime. In a fully disclosed market, the behaviour of institutional investors is usually subject to a strong imitation effect (BHS 1995, Duong et al., 2009). Conversely, the informativeness of a trading strategy is obviously weaker in the case of hidden identities. This research looks for a confirmation of the preceding results on order aggressiveness, and it follows Foucault et al. (2007) who examine the LOB post market reform and find a decrease of the bid-ask spread informativeness in the new setting. Furthermore, our data set allows for a comparison with studies on non-anonymous markets, as Cao et al. (2008) who examine the Australian Stock Exchange where institutional investors have full access to the identities of traders. A second point concerns the categories of traders operating at Euronext Paris. The Paris Bourse presents some intermediaries called 'animateurs' who enhance the market liquidity by assuring a maximum spread size and a minimum level of depth. Aitken et al. (2007) study the role of these operators at Euronext Paris and find a limited contribution to the overall liquidity of the LOB. Although Euronext Paris may be mainly regarded as an order-driven market, the presence of agents who assure contractual levels of liquidity is a challenge with respect to the results obtained in *pure* order-driven markets (see Ranaldo, 2004 or Cao et al., 2009). Finally, the last point regards the traditional approach to study order aggressiveness. As a matter of fact, most of the previous empirical contributions simply apply the scheme in BHS (1995) by means of discrete choice models. This strategy is particularly appealing, but we show that it disregards relevant quantitative differences within the same order category. Therefore, we still adopt the BHS's (1995) classification that represents the leading reference in this topic, but we propose an alternative approach to capture the heterogeneity of orders and to fully exploit the informativeness of price and volume.

This paper is organized as follows. Section 5.2 reviews the original BHS's (1995) ranking and it introduces our approach to measure order aggressiveness. Section 5.3 presents the data set used in this paper and provides some descriptive statistics for order aggressiveness. Section 5.4 describes the model adopted for the empirical analysis and discusses the microstructure hypotheses. Section 5.5 presents the empirical results and examines the price impact for different categories of orders. Finally, Section 5.6 concludes.

5.2 Order Aggressiveness

5.2.1 Ranking order aggressiveness

Most of the empirical papers on order aggressiveness use or slightly adapts the scheme in BHS (1995) to measure traders' impatience. BHS (1995) essentially rank each order according to its price position with respect to the best quotes. For an agent entering the market as a buyer, BHS (1995) sort the aggressiveness of his order as follows:

1. **Large buy**: when the agent specifies a limit price above BASK[1] and demands a quantity larger than that available at BASK;

2. **Market buy**: is a market order demanding a quantity larger than that available at BASK;

3. **Small buy**: when the agent specifies a limit price above BASK, but demands a quantity smaller than that available at BASK;

4. **Limit order within the quotes**: when the agent submits a limit order with a price improving BBID, but that is under BASK;

5. **Limit order at the quote**: when the buyer specifies a limit order with a price at BBID;

6. **Limit order below the quote**: when the trader submits a buy limit order with a price below BBID;

The approach can be easily extended to the case of sell orders, by considering the symmetrical ranking. The BHS's (1995) scheme splits all the orders into two broad groups: classes 1-3 include orders that result in an immediate matching and subtract liquidity from the market (market orders and marketable limit orders). On the other hand, the last three classes collect less aggressive orders that enhance market liquidity and do not

[1]To simplify the exposition, in the following we refer to the two prevailing quotes as BASK (best ask) and BBID (best bid).

imply an instantaneous execution. From the ranking, it is immediate to recover the two components that determine the level of aggressiveness: the price and the volume. Clearly, the price of an order represents the leading indicator in markets ruled under a price priority system; this is also suggested by the fact that the volume is a discriminant factor only for the first three categories. Nevertheless, which is coherent with the previous literature emphasizing the role of volume for order aggressiveness (Lo and Sapp, 2010), we differentiate between *price aggressiveness* and *volume aggressiveness*.

5.2.2 An alternative approach

The simplicity of BHS's (1995) classification and the necessity to deal with price discreteness explain why most of the previous contributions have investigated order aggressiveness using a categorical variable approach. However, despite the widespread diffusion of this method, the application of discrete variable models does not seem to be a binding choice for our sample where the tick size equals to 0.005 Euro. As Aitken et al. (2005) highlight for the Australian Stock Exchange, such a tick size makes the price discreteness issue less relevant for the study of market microstructure hypotheses. This is not a secondary point, as BHS's (1995) ranking is essentially based on a *quasi*-qualitative criterion. Indeed, BHS (1995) do not categorize orders on the basis of a quantitative scale, and this inexorably collapses the informativeness of both price and volume. To realize this point for price aggressiveness, it is sufficient to consider the first class, assuming that BASK is 50 Euros and the quantity available is 200 shares. If two buyers submit two limit orders for 500 shares at a price of 50.005 and 53 Euros respectively, both orders will be assigned to the same class according to BHS (1995). However, it seems quite questionable to associate the same level of price aggressiveness to the two orders, as the second trader is disposed to buy at a price that is considerably higher than the first limit price. Therefore, we propose to measure price aggressiveness as the reservation price of a trader, and we introduce the following variable Δp_t:

$$\Delta p_t = \begin{cases} BASK_t - p_t & \text{for buy orders} \\ \\ p_t - BBID_t & \text{for sell orders} \end{cases} \tag{5.1}$$

where p_t is the price of an order submitted at time t, and $BASK_t$ and $BBID_t$ represent the prevailing quotes at time t. In this way, price aggressiveness is expressed as a continuous variable that measures the distance from the best quote on the opposite side of the market. According to the type of order, Δp_t assumes the following values:

$$\Delta p_t \begin{cases} < 0 & \text{for marketable limit orders} \\ \\ = 0 & \text{for market orders} \\ \\ > 0 & \text{for limit orders.} \end{cases} \tag{5.2}$$

Regardless of the quantity, Δp_t assumes a negative value for buy (sell) orders displaying a limit price higher (lower) than BASK (BBID). On the other hand, Δp_t is equal to zero when the traders submit a market order, and it assumes a positive value for limit orders that belong to the classes 4 to 6 of BHS's (1995) ranking. By means of Δp_t, we measure the price aggressiveness under an ex-ante perspective, i.e. as a sort of willingness to pay. Generally speaking, it is the relative price position which matters to evaluate the price aggressiveness of an order and the corresponding probability of execution. When a trader submits a buy order, his level of aggressiveness will be directly related to how much he is likely to pay over the standing BASK, and similarly for a sell order. Then, the likelihood of execution progressively decreases for orders submitted at less favourable price conditions: for a buy order this corresponds to a price lower than BASK, for a sell order this is consistent with a price higher than BBID.

With respect to BHS's (1995) ranking, our approach diminishes the number of order categories and it also collapses the marketable limit orders in a unique class. Now, some considerations are in order to motivate this choice. First of all, the lower number of classes is actually compensated by the quantitative measurement of price aggressiveness and Section 5.3.2 offers a comprehensive discussion on this point. Second, though BHS (1995) represents the primary reference in this field, several papers adopt some modifications of it. As an example, Duong et al. (2009) propose a classification which is fairly similar to the one implied by Δp_t, but they still adopt a categorical variable model. Third, even though we collect all the marketable limit orders in a unique category, we assess the impact of volume via a distinct equation, in line with Lo and Sapp (2010). Finally, our approach seems more coherent with Euronext trading rules. In fact, the traders are aware that a marketable limit order incurs in less favourable prices when it walks down (up) the book to be fully matched. Conversely, in the case of market orders, Euronext sets a highly narrow price collar which prevents to walk far away from the best quotes. Therefore, the price tolerance that a trader is willing to accept seems a better measurement of the level of aggressiveness, especially when the LOB is not deep or when the traders submit large quantities.

As it concerns volume aggressiveness, we recognize that the order size is not explicitly modelled in BHS (1995), and similarly in most of the following empirical contributions. This choice seems coherent with market priority rules, but it disregards the informational content of quantity. We believe that a separate measurement of volume represents a more suitable alternative, rather than simply considering a few discriminant thresholds. This is particularly true for aggressive orders, and to figure out this point we re-consider the previous example with a BASK of 50 Euros and a standing volume of 200 shares. Then, we assume two incoming buy orders with a limit price equal to 50.1 Euros and a quantity which is 300 and 5000 shares, respectively. By neglecting the quantitative impact of volume, we would assign the two orders to the same ranking class, though the association would be fairly questionable also in this case. Such an issue is even more problematic in the case of market orders where the volume represents the only condition which is set by traders. All this suggests that a single dimension is not sufficient to classify order aggressiveness; therefore, we adapt the model in Lo and Sapp (2010) to simultaneously study the two principal terms of an order submission strategy.

5.3 Data

5.3.1 The sample

The data set used in this paper is provided by Eurofidai and it includes all the orders submitted to Euronext Paris LOB[2], between 1 September 2009 and 30 October 2009, for six stocks of the CAC 40 index, namely: Alstom, Axa, Crédit Agricole, Eads, Essilor, and Sanofi-Aventis. We single out three main categories of orders from the data set:

1. The **marketable limit orders**, which are limit orders with a price improving the best quotes; these orders are promptly matched and fully executed as long as price conditions are satisfied.

2. The **market orders**, which are immediately filled at the best quotes standing on the market. If the quantity available at the best quotes is not sufficient, the order is fulfilled at different prices, but the traded price has to fit within highly narrow collars set by the Exchange.

3. The **limit orders** wherein the price does not improve the best quotes; these orders fill the LOB and increase the market liquidity without resulting in a trade.

Similar to Ranaldo (2004), the data set only supplies information on the prevailing quotes, and no data are available for the rest of the LOB. Empirical microstructure literature has long discussed the LOB informativeness outside the best quotes, and a final answer is yet to be given. For instance, Cao et al. (2008) find a limited informational content for the lower levels of the LOB, while Lo and Sapp (2010) find mixed results for the informativeness of quotes behind the best prices, and Pascual and Veredas (2009) show that the model goodness-of-fit increases by considering the whole LOB. Nevertheless, most

[2]There is no information on hidden orders. For the empirical analysis, we only consider orders submitted during the continuous auction, i.e. between 9.00 a.m. and 5.30 p.m..

Table 5.1: Descriptive statistics for orders submitted to Euronext LOB between 1 September and 30 October 2009.

	Volume	Ask	Bid	Midpoint	Spread	Volatility	Time	Observations
Alstom	562.2	494.9	482.8	49.80	0.035	0.0094	0.944	1,423,163
Axa	1212.4	2114.5	2029.2	17.85	0.012	0.0031	0.644	2,090,483
Crédit A.	1453.2	2365.8	2237.6	13.89	0.012	0.0034	0.911	1,477,359
Eads	1271.2	1796.1	1732.8	14.86	0.019	0.0036	1.396	963,960
Essilor	362.4	594.2	589.2	38.62	0.030	0.0058	1.759	764,576
Sanofi-A.	560.5	799.2	774.0	49.64	0.022	0.0052	0.545	2,468,192

Table 5.2: Descriptive statistics for order size with respect to Δp_t intervals.

	Alstom					
	Buy			Sell		
Volume	$\Delta p_t < 0$	$\Delta p_t = 0$	$\Delta p_t > 0$	$\Delta p_t < 0$	$\Delta p_t = 0$	$\Delta p_t > 0$
0.25 perc.	90	84	200	100	81	200
0.5 perc.	177	173	216	190	171	204
0.75 perc.	318	325	400	345	318	400
Mean	307.01	276.71	582.69	329.13	267.76	607.58
S.D.	547.31	430.63	3,267.41	599.61	415.74	3,384.31
Relat. frequency	0.023	0.084	0.893	0.022	0.082	0.896

	Eads					
	Buy			Sell		
Volume	$\Delta p_t < 0$	$\Delta p_t = 0$	$\Delta p_t > 0$	$\Delta p_t < 0$	$\Delta p_t = 0$	$\Delta p_t > 0$
0.25 perc.	109	199	300	119	163	300
0.5 perc.	310	452	500	348	420	500
0.75 perc.	771	900	898	810	863	900
Mean	674.61	707.49	1,295.20	758.91	695.77	1,383.17
S.D.	1,257.09	921.07	6,644.63	1,812.47	955.97	7,001.68
Relat. frequency	0.018	0.087	0.895	0.018	0.091	0.891

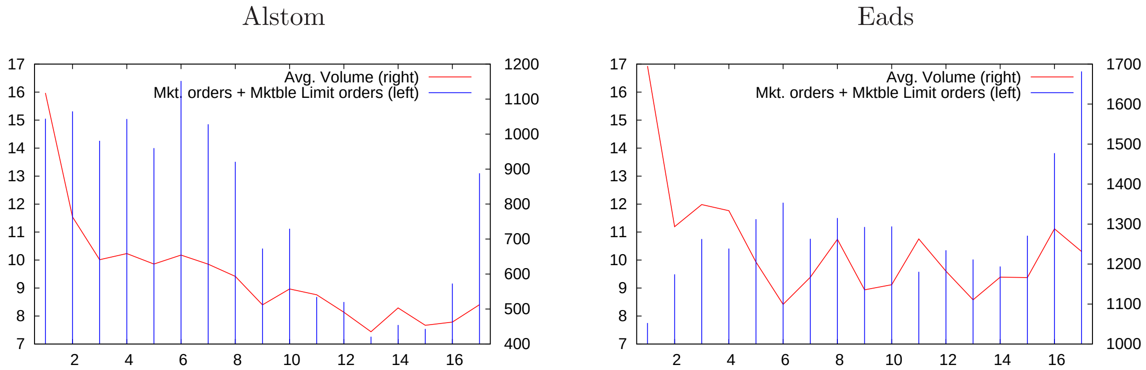
of the literature unanimously recognize the two best quotes as the principal repository of the market information, and our measure of price aggressiveness is coherent with this hypothesis. We exclude from the original data set the orders registered as applications (trades internally executed by financial institutions) and the marketable limit orders with a price improving the best quote by more than 5 Euros. As a whole, this corresponds to a very marginal part of the observations (always less than 0.5 %), while it contributes to minimize the inclusion of outliers due to order processing.

Table 5.1 presents some descriptive statistics. The first three columns exhibit the average order volume and the average quantity available at the best quotes. The fourth and the fifth columns report the average midpoint and the average absolute spread, respectively. The sixth column displays the 5-minute standard deviation computed on the midpoint, the seventh column provides the inter-order duration expressed in seconds, while the last column shows the number of observations. Table 5.2 reports some descriptive statistics for the order size with respect to Δp_t intervals, and it also indicates the frequency of each category of orders. For the sake of brevity, we limit the presentation only to Alstom and Eads, but our findings extend to the rest of the sample and are available upon request. According to Table 5.2, around 10% of the orders submitted to the LOB result in a trade; this percentage is a bit lower than the one reported in previous empirical studies of Euronext Paris (e.g. De Winne and D'Hond, 2005). Table 5.2 also indicates that limit orders are usually larger than marketable limit orders or market orders; this result is valid for the whole sample and it is in line with previous contributions, see Harris and Hasbrouck (1996), Bae et al. (2003), Lo and Sapp (2010). The inverse relationship between price aggressiveness and order size is coherent with the hypothesis of volume informativeness (Easley and O'Hara, 1987). As a matter of fact, the traders internalize that large aggressive orders are more revealing of a private signal, and they place a smaller volume for market orders and marketable limit orders. Finally, Figure 5.1 exhibits the intraday seasonality pattern for price and volume aggressiveness; both plots only take into account market orders and marketable limit orders. The figure display the usual U-shaped pattern for the average volume size. On the other hand, the trend of order submission differs between Alstom and Eads, even though both stocks display a remarkable intraday periodicity with respect to Δp_t.

5.3.2 Heterogeneity in order aggressiveness

We examine the degree of heterogeneity for orders that belong to the same class according to the BHS's (1995) ranking and we find that they are not always collapsed in uniform

Figure 5.1: Intraday plots for Δp_t and order size. The plots report the percentage of market and marketable limit orders together with the corresponding average volume. The x-axis provides the 30-minute intervals of the continuous auction.



categories. In this context, a quantitative measurement of order aggressiveness is proven to be a preferable strategy. Tables 5.3 to 5.8 provide the contingency tables for classes of order aggressiveness with respect to Δp_t and order size. The tables follow the sequence of BHS's (1995) ranking and, to save space, we limit to buy orders for Alstom and Eads; nevertheless, our findings also extend to the remaining stocks and to the sell side of the market as well. Any multiple of the tick size away from the best quote is considered as a distinct level of traders' impatience, in line with the discussion in Section 5.2. Our analysis is similar to Cao et al. (2008) who classify the position of orders in the LOB in terms of steps. We adopt the tick size as the step to determine the distance of an order from the best quotes, and we sort the values of Δp_t into five intervals. The first interval includes orders that are just one tick away from the best quotes. The second collects orders that are between two and five ticks away; the third has between six and ten ticks, the fourth between eleven and fifteen ticks, while the fifth interval includes all the orders that are more than fifteen ticks away from the best quotes. The volume is also sorted into five intervals that classify the order size on the basis of the number of shares.

Table 5.3 refers to marketable limit orders demanding a volume larger than that available at BASK, and it can be compared with Table 5.5 that considers fully executed limit orders improving BASK. As to price aggressiveness, both tables display a remarkable portion of observations in the first interval, i.e. just one tick above BASK. This outcome is not unexpected, since the data set consists of extremely liquid stocks and a relevant percentage of transactions generally occurs around the best quotes. The result is particularly evident for Eads which exhibits a lower average price; thus, less ticks are required to generate a specific price improvement in relative terms. However, and independently of the order size, Table 5.3 shows that around 80 (50) percent of the observations for Alstom (Eads) improves BASK by more than one tick; the proportion is even higher in Table 5.5 and it equals to 85 (65) percent for Alstom (Eads). The frequency of the observations clearly diminishes in extreme Δp_t intervals, though some remarkable percentages are still evident, especially in Table 5.5 where around 40 (20) percent of Alstom (Eads) orders are placed at more than five ticks from BASK[3]. According to both tables, a relevant amount of price informativeness would be lost by limiting to a threshold criterion to discriminate the observations, as in BHS (1995). Conversely, with Δp_t we quantitatively evaluate the maximum level of price tolerance that a trader is willing to accept. This seems more suitable when there is price heterogeneity within a class or when the orders encounter a multi-price execution by walking down (up) the book. As it concerns the order size, Tables 5.3 and 5.5 exhibit a wide dispersion among all the classes, with the only exception of extremely large orders (>5000); the effect is particularly evident for Eads and for less aggressive submissions ($\Delta p_t > -0.05$). These figures confirm the existence of a fair volume heterogeneity, and they seem coherent with our simultaneous model for the price-quantity decision.

Table 5.4 displays the marginal frequencies for volume intervals in the case of market orders. We limit to volume aggressiveness as Δp_t always equals to zero for market orders[4]; therefore, the degree of heterogeneity for market orders is essentially measured by the

[3]The results for small buys show that a higher frequency of observations is placed in the more negative intervals of Δp_t. Although an in-depth analysis on this point is out of the scope of this research, our findings may be traced back to volume informativeness. The traders, especially the informed ones, do not have any incentive to combine price aggressive orders with a large order size, as it could be revealing of their private information. A contribution on this topic is provided by the stealth trading literature, see Chakravarty (2001).

[4]Though a market order can walk the book to be completely filled, the price tolerance allowed by Euronext is very limited and these orders are generally matched around the best quotes, especially for highly liquid stocks. This matches with the choice to set $\Delta p_t = 0$ for market orders, also because we measure the ex-ante price tolerance and not the ex-post execution price.

Table 5.3: Contingency table for Volume and Δp_t intervals. The frequencies refer to *Large buys* for Alstom and Eads.

	Alstom Δp_t						Eads Δp_t					
Volume	[-5;-0.08]	[-0.075;-0.055]	[-0.05;-0.03]	[-0.025;-0.01]	[-0.005]	Total	[-5;-0.08]	[-0.075;-0.055]	[-0.05;-0.03]	[-0.025;-0.01]	[-0.005]	Total
[1;100]	2.14	0.92	3.63	8.38	3.00	**18.07**	0.03	0.15	1.52	4.91	3.65	**10.26**
]100;500]	2.47	2.55	8.77	33.82	13.20	**60.80**	0.31	0.54	3.01	14.33	17.31	**35.50**
]500;1000]	1.28	0.80	1.94	7.48	2.43	**13.93**	0.62	0.31	1.41	8.67	13.79	**24.80**
]1000;5000]	1.47	0.47	1.07	2.97	0.94	**6.92**	0.72	0.57	2.68	10.21	12.97	**27.14**
>5000	0.14	0.02	0.01	0.07	0.04	**0.28**	0.10	0.08	0.51	1.00	0.59	**2.29**
Total	**7.50**	**4.76**	**15.42**	**52.72**	**19.61**	**100.00**	**1.78**	**1.65**	**9.13**	**39.12**	**48.31**	**100.00**

Table 5.4: Marginal frequencies of Volume for Alstom and Eads. The frequencies refer to *Market buys*.

	Alstom	Eads
Volume	Realtive Frequency	Relative Frequency
[1;100]	32.36	15.96
]100;500]	55.05	39.54
]500;1000]	9.16	24.47
]1000;5000]	3.37	19.45
>5000	0.05	0.58
Total	**100.00**	**100.00**

Table 5.5: Contingency table for Volume and Δp_t intervals for Alstom and Eads. The frequencies refer to *Small buys*.

	Alstom Δp_t						Eads Δp_t					
Volume	[-5;-0.08]	[-0.075;-0.055]	[-0.05;-0.03]	[-0.025;-0.01]	[-0.005]	Total	[-5;-0.08]	[-0.075;-0.055]	[-0.05;-0.03]	[-0.025;-0.01]	[-0.005]	Total
[1;100]	14.04	4.43	9.21	19.95	5.18	**52.80**	2.97	2.54	3.87	15.53	9.57	**34.47**
]100;500]	3.85	2.04	5.50	23.26	8.20	**42.85**	2.21	1.84	2.93	18.96	17.04	**42.98**
]500;1000]	0.25	0.10	0.22	2.23	0.68	**3.49**	0.43	0.16	1.06	7.51	7.02	**16.18**
]1000;5000]	0.01	0.03	0.06	0.59	0.15	**0.85**	0.14	0.23	0.59	3.13	2.19	**6.28**
>5000	0.00	0.00	0.00	0.01	0.00	**0.01**	0.02	0.00	0.00	0.00	0.06	**0.08**
Total	**18.15**	**6.60**	**14.99**	**46.04**	**14.21**	**100.00**	**5.77**	**4.77**	**8.45**	**45.13**	**35.88**	**100.00**

Table 5.6: Contingency table for Volume and Δp_t intervals for Alstom and Eads. The frequencies refer to *Limit (buy) orders within* the best bid and the best ask.

	Alstom Δp_t						Eads Δp_t					
Volume	[0.005]	[0.010;0.025]	[0.03;0.05]	[0.055;0.075]	[0.08;5]	Total	[0.005]	[0.010;0.025]	[0.03;0.05]	[0.055;0.075]	[0.08;5]	Total
[1;100]	2.94	11.18	5.69	1.07	0.30	**21.17**	7.29	3.73	0.43	0.09	0.00	**11.53**
]100;500]	7.29	36.02	23.18	4.16	0.71	**71.36**	25.11	16.17	4.02	0.46	0.12	**45.88**
]500;1000]	0.77	3.22	1.16	0.52	0.16	**5.84**	14.90	7.37	1.87	0.14	0.05	**24.33**
]1000;5000]	0.23	0.98	0.23	0.09	0.07	**1.59**	12.39	4.36	1.06	0.07	0.02	**17.89**
>5000	0.00	0.02	0.00	0.00	0.01	**0.03**	0.22	0.12	0.03	0.00	0.00	**0.37**
Total	**11.23**	**51.42**	**30.26**	**5.84**	**1.25**	**100.00**	**59.91**	**31.75**	**7.41**	**0.76**	**0.19**	**100.00**

Table 5.7: Contingency table for Volume and Δp_t intervals for Alstom and Eads. The frequencies refer to *Limit (buy) orders at* the best bid.

	Alstom Δp_t						Eads Δp_t					
Volume	[0.005]	[0.010;0.025]	[0.03;0.05]	[0.055;0.075]	[0.08;5]	Total	[0.005]	[0.010;0.025]	[0.03;0.05]	[0.055;0.075]	[0.08;5]	Total
[1;100]	0.85	8.67	9.10	3.01	1.13	**22.76**	2.43	4.92	1.41	0.64	0.64	**10.03**
]100;500]	2.20	18.64	25.20	16.75	5.28	**68.07**	9.86	23.31	14.12	1.40	0.86	**49.54**
]500;1000]	0.41	1.78	1.51	1.05	1.26	**6.01**	5.28	11.17	7.13	1.19	0.98	**25.76**
]1000;5000]	0.25	1.14	0.60	0.27	0.67	**2.93**	3.38	5.65	2.45	1.28	1.00	**13.75**
>5000	0.02	0.12	0.05	0.03	0.01	**0.23**	0.44	0.26	0.07	0.03	0.12	**0.91**
Total	**3.73**	**30.35**	**36.46**	**21.11**	**8.35**	**100.00**	**21.39**	**45.31**	**25.18**	**4.54**	**3.60**	**100.00**

Table 5.8: Contingency table for Volume and Δp_t intervals for Alstom and Eads. The frequencies refer to *Limit (buy) orders below* the best bid.

	Alstom Δp_t						Eads Δp_t					
Volume	[0.005]	[0.010;0.025]	[0.03;0.05]	[0.055;0.075]	[0.08;5]	Total	[0.005]	[0.010;0.025]	[0.03;0.05]	[0.055;0.075]	[0.08;5]	Total
[1;100]	0.87	0.50	2.39	2.53	3.35	**9.63**	0.34	0.57	0.82	0.78	1.56	**4.07**
]100;500]	0.55	2.66	21.23	30.01	15.71	**70.16**	0.39	15.95	24.99	3.40	2.08	**46.81**
]500;1000]	0.13	0.16	1.05	2.33	11.31	**14.98**	0.16	6.05	15.78	3.81	1.81	**27.61**
]1000;5000]	0.05	0.03	0.09	0.17	3.70	**4.04**	0.13	2.03	8.33	3.00	5.18	**18.67**
>5000	0.00	0.00	0.00	0.01	1.18	**1.19**	0.01	0.02	0.08	0.13	2.60	**2.85**
Total	**1.60**	**3.35**	**24.76**	**35.05**	**35.25**	**100.00**	**1.03**	**24.62**	**50.00**	**11.12**	**13.23**	**100.00**

amount of shares requested to the market. Also in this case, extremely large orders represent a small proportion of the observations, while there exists a fair heterogeneity for the remaining classes. A plain application of BHS (1995) ranking would imply a unique classification, irrespective of the number of shares of each order. Conversely, the approach suggested by Lo and Sapp (2010) seems more appropriate, especially when the volume represents the main indicator to discriminate the aggressiveness of traders.

Tables 5.6 to 5.8 provide joint and marginal frequencies for orders that do not immediately result in a transaction. For these classes, Δp_t is always strictly positive, with larger values corresponding to lower levels of impatience. Table 5.6 offers the descriptive statistics for buy limit orders which improve BBID but are below BASK. For limit orders submitted within the quotes or directly at the quotes (see Table 5.7), Δp_t is proportional to the spread width and it measures the minimum price improvement required to match an order. In Table 5.6, we notice the high concentration of orders within few ticks from BASK, especially for Eads. This finding is not striking as the stocks of the CAC 40 index comply with high liquidity standards and it would be unlikely to observe a large spread and, consequently, a high value of Δp_t. Nevertheless, around 35 (8) percent of Alstom (Eads) orders are placed in the three most positive intervals of Δp_t. Therefore, our approach seems particularly appropriate during periods of large spreads, when the submission of orders within the quotes does not imply a higher probability of execution *per se*. With regard to volume heterogeneity, Table 5.6 exhibits a certain level of data dispersion, which is still more apparent for Eads. Similar to market orders, this again emphasizes the role of volume to fully describe order aggressiveness when the price informational content is limited.

Table 5.7 displays the descriptive statistics for orders submitted at BBID. In this case, Δp_t intervals provide an indirect measure of the bid-ask spread and, coherent with the previous discussion, it is not surprising that a relevant concentration of orders is submitted just a few ticks away from BASK. Nevertheless, more than 65 (35) percent of Alstom (Eads) orders are placed at more than five ticks from BASK. As to volume, Table

5.7 still exhibits a noticeable variability, despite the dominant fraction of orders between 100 and 500 shares and the marginal presence of extremely large orders.

Frequency statistics for buy limit orders with price below BBID are collected in Table 5.8. The table presents a remarkable variability for price and volume aggressiveness, with a striking percentage of orders placed at higher values of Δp_t. On the basis of BHS's (1995) ranking, we should collapse both the few orders submitted close to BASK and the high percentage of orders placed at more than 15 ticks from BASK into the same class, though the latter are clearly less likely to be executed. According to Table 5.8, the traders also provide liquidity far behind from the best quotes, and this feature would not be fully captured by using discrete variable models. Finally, the quantity is also fairly distributed among the five classes, though the range]100;500] is still dominant.

5.4 Model and Empirical Specification

5.4.1 The model

We separately examine the two components of order aggressiveness and we model the volume and the price of each order by means of a simultaneous equation system

$$\Delta p_t = f_p(x_t;\theta_p) + \phi q_t + \epsilon_1 \tag{5.3}$$

$$q_t = f_q(x_t;\theta_q) + \gamma \Delta p_t + \epsilon_2 \tag{5.4}$$

$$\rho = Corr(\epsilon_1, \epsilon_2)$$

where $f_p(x_t;\theta_p)$ is a function of a set of regressors x_t with associated parameters θ_p, and an analogous specification holds for the volume of each order, q_t. The parameter ρ allows for the presence of correlation between the two error components, ϵ_1 and ϵ_2. This setting slightly differs from the one proposed by Lo and Sapp (2010) who consider an ordered probit for price aggressiveness and a censored regression for the quantity. In the previous sections, we have widely discussed the reasons to exclude a discrete variable model for price aggressiveness. As to quantity, Lo and Sapp (2010) employ a censored regression model to account for the clustering of volume which characterizes their data set; however, no empirical evidence suggests to use the same approach for our sample.

The endogeneity that affects the price-quantity decision is solved through a two-stage least squares estimation where the lagged dependent variables are employed as instruments. As emphasized by Lo and Sapp (2010), the endogeneity between price and volume has not just an econometric implication, but it also relates to the two fundamental components of the traders' submission strategies. In fact, when a trader chooses the price and the quantity of his order, he faces a trade-off between instantaneous matching and costs of execution. On the one hand, he ensures a prompt execution by submitting price aggressive orders. On the other hand, he exposes to private information disclosure or to unfavourable price execution when his order has to walk down the book to be fully matched. As a solution, the traders generally combine price aggressiveness with the splitting of sizeable orders into smaller quantities.

5.4.2 Explanatory variables and microstructure hypotheses

For the empirical analysis, we consider a wide set of explanatory variables that define the status of the LOB and that may influence the aggressiveness of traders. These variables are included as regressors in the system (5.3)-(5.4) and may be thought as the investor information set (Beber and Caglio, 2005). The full list of variables includes:

- **Bid-ask spread**. It is defined as the difference between the standing BASK and BBID at the time of order submission. The width of the spread is affected by multiple factors, as informational asymmetries or inventory costs, but the first component is clearly dominant in an order-driven market. The spread measures the gap between the two sides of the market and it is inversely related to the level of competition and liquidity. As a matter of fact, the spread widens in case of less liquid stocks or when market participants perceive an increasing likelihood of informed trading (see Glosten and Milgrom, 1985, Easley and O'Hara, 1987). In this context, the uninformed traders are exposed to the risk of being picked off by informed agents and the submission of market orders becomes a costly strategy, as the traders would pay the whole distance between the best quotes. Coherent with previous empirical studies (e.g. Biais et al., 1995, Bae et al., 2003, and Ranaldo, 2004), we expect the spread to be negatively related to price aggressiveness, with traders being less impatient when the spread increases. As to quantity, when the spread is high the trader prefer to submit smaller orders to minimize the loss in case of trading with informed agents. On the other hand, the traders are also encouraged to gain from the provision of market liquidity through the submission of large limit orders. Despite of this opposite pressure, Lo and Sapp (2010) find a negative relationship between order size and spread, though their estimates are not always significant.

- **Same side depth**. For each buy (sell) order, the depth on the same market side is measured as the number of shares standing at BBID (BASK). Parlour (1998) proposes a model where the probability of execution of each order depends on the future order arrivals and on the volume available in the LOB. Unless of a price improvement, rational agents know that submitting a limit order on a thick market increases the execution risk because of time priority rules. When a considerable amount of volume is available at BBID (BASK), the buyer (seller) has to force price aggressiveness in order to jump over the queue of preceding orders. The competition among traders in terms of price improvement and the resulting submission of aggressive orders is referred to as the crowding-out effect. Generally speaking, the greater the depth at BBID (BASK), the higher is the aggressiveness of an incoming buyer (seller). Therefore, we expect to find a positive relationship between price aggressiveness and volume available at the same side of the market, as in Ranaldo (2004) or Pascual and Veredas (2009). As to volume aggressiveness, we expect a negative relationship between order size and depth on the own side, since the traders seeking for a faster execution combine the high price aggressiveness with a small quantity (see Lo and Sapp, 2010 or Hall and Hautsch, 2006).

- **Opposite side depth**. It is measured as the number of shares available at BASK (BBID) for a buy (sell) order, similar to the previous regressor. When the buy (sell) side is thick, the execution risk for sellers (buyers) is reduced and the traders are not obliged to implement price aggressive strategies to execute their orders. In Parlour (1998), this is known as the strategic effect and we expect a negative relationship between the opposite side depth and the price aggressiveness of an incoming order. Conversely, we expect a positive impact on volume aggressiveness, as a high depth on the opposite side of the market reduces the execution risk of submitting a large quantity.

- **Volatility**. We measure price volatility as the 5-minute standard deviation of midquote log returns, partially following Ranaldo (2004). The relationship between volatility and order submission has received a large attention in the literature, see Handa and Schwartz (1996), Foucault (1999), and Foucault et al. (2005). In general, price volatility has two main sources, i.e. liquidity trading and information asymmetries; however, the traders are clearly more concerned with the latter, as they may face agents with superior information. Therefore, in the case of information-driven volatility we expect a positive impact of volatility on Δp_t, as the traders submit less aggressive orders to protect themselves from the pick-off risk. The effect of volatility on quantity is examined by Ahn et al. (2001), Bae et al. (2003), Hasbrouck and Saar (2002), and Lo and Sapp (2010), among others. The previous contributions provide evidence of a mixed result for this regressor, which is found both positively and negatively associated with liquidity.

- **Return**. It is defined as the 5-minute-midquote log return, similar to Ellul et al. (2003). Technical analysts look at stock returns as a signal of market direction: a positive return is considered as a proxy for good news, while negative returns are associated with bad news. A trader seeking for momentum strategies submits price aggressive orders in line with the market direction to profit from short-lived investment opportunities. Clearly, stock returns have an asymmetrical impact on the two sides of the market. For the buy side, we expect to observe a positive relationship between returns and price aggressiveness, while the contrary holds for the sell side. On the other hand, the effect for volume aggressiveness is expected to be the reverse of that described for Δp_t. In fact, technical traders combine price aggressiveness with small quantity to ensure a fast execution. In case of positive returns, this implies a negative impact for the volume of buy orders, while the opposite holds for sell orders.

- **Time**. Time is measured as the number of seconds between two consecutive orders. Easley and O'Hara (1992) attribute an informational content to time intervals, such that a high frequency of transactions is consistent with an increasing presence of informed agents. Actually, a faster order submission process is perceived as a trustworthy signal of the presence of information asymmetries; these asymmetries lead to a widening of the spread and discourage the submission of price aggressive orders. Coherent with Easley and O'Hara (1992), we expect the time to be positively correlated with price aggressiveness. As to volume, the frequency of order submission should exert a negative effect on the order size, in line with the previous discussion for the bid-ask spread.

- **Temporal patterns**. The presence of a daily periodicity in order submission has been widely documented since BHS (1995) seminal paper, and it is also fairly evident from Figure 5.1. According to Bloomfield et al. (2005), the informed traders submit aggressive orders at the start of the continuous trading session to profit from the uncertainty which characterizes the market opening. Subsequently, price discovery and public information disclosure are expected to reduce price aggressiveness throughout the day. At the end of the day, temporal constraints and the arrival of new information push the traders to close their open positions, and a new rising in price aggressiveness is more likely to occur (BHS, 1995 or Lo and Sapp, 2010). To account for the presence of daily patterns in price aggressiveness, we consider a specification for time-of-day effects which replicates the one in Ferriani (2012). First of all, we model the daily seasonality by means of the variable δ_k evolving from zero to one throughout the continuous auction. Secondly, we try to single out an intraday pattern by means of four dummies that identify some critical periods during the continuous auction, namely D_{open}, D_{lunch}, D_{SP}, and D_{clos}. The variable D_{open} is equal to one for orders submitted between 9.00 A.M. and 9.30 A.M., D_{lunch} between 00.00 P.M. and 02.00 P.M., D_{SP} between 3.00 P.M. and 4.30 P.M., and finally D_{clos} between 5.00 P.M. and 5.30 P.M.. We include these intervals to highlight some stylized facts of a traditional trading day. The opening and the closing periods usually display a higher frequency of transactions as well as a remarkable proportion of institutional trading. The lunch period is characterized by a strong decrease in the trading activity, while D_{SP} coincides with the opening of the U.S. Stock Exchange and it singles out a period of market pressure. All these variables are also included in the equation for quantity, given the broad literature that documents the existence of intraday seasonality in the order volume. Volume aggressiveness is expected to follow an inverse path with respect to Δp_t, since it is inconvenient to submit large price aggressive orders because of volume informativeness; nevertheless, the graphical analysis of Figure 5.1 shows that this is not always the case.

- **Lagged dependent variable**. The existence of a strong autocorrelation pattern in the order submission process has been broadly discussed in the literature, as in BHS (1995), Al-Suhaibani and Kryzanowski (2000), and Pascual and Veredas (2009). BHS (1995) explain this serial correlation by referring to the strategic order splitting and to the imitation effect that characterizes the investment strategies of traders. The autocorrelation pattern involves both price and volume aggressiveness and we expect a positive relationship between current and previous submissions, both in terms of price and volume.

- **Signed cumulative volume**. This variable is computed as the 5-minute signed cumulative volume, according to the following expression:

$$Sigcum_t = \sum_{5\,min} sign_t \sqrt{Volume_t} \tag{5.5}$$

where $sign_t$ is a variable that identifies the market direction and it equals to +1 for buy orders and -1 for sell orders. $Sigcum_t$ evaluates the impact of a directional demand for liquidity arising from the trading pressure on a specific side of the market. The signed cumulative volume has been employed by Hasbrouck (2009), Goyenko et al. (2009), and Cao et al. (2009) to measure the price impact of trades. To be consistent with their analysis, we only consider market orders and marketable limit orders, i.e. the orders that effectively result in a trade. In terms of price aggressiveness, when $Sigcum_t$ increases (decreases), the market participants perceive a bullish (bearish) tendency, and are willing to submit more aggressive orders on the buy (sell) side. This is similar to the momentum-type effect described for the stock return. Consequently, the volume is expected to follow the inverse pattern, with price aggressive orders associated with a small quantity.

- **Book imbalance**. It is defined as the difference between the amounts of shares available at the best quotes, expressed in absolute value. Book imbalance does not take into account the market direction, it simply augments whenever there exists a disequilibrium in the depth between BBID and BASK. In terms of price aggressiveness, Parlour (1998) shows that an increasing book imbalance should encourage the traders to submit less aggressive orders on the thin side and more aggressive orders on the thick side; then, the overall impact on Δp_t should reflect the dominant pressure between the two. As to quantity, we expect to find an inverse relationship with price aggressiveness, on the basis of the previous discussion for depth at the best quotes.

5.5 Empirical results

5.5.1 The simultaneous equation model

This section discusses the results of the simultaneous equation model introduced in Section 5.4.1. Table 5.9 displays the parameter estimates and the p-values for price aggressiveness, while Table 5.10 reports the same statistics for quantity. We estimate the model for buy and sell orders separately, as it is standard in the literature on order aggressiveness. Because of the relevant sample size, we base the following analysis on a 1% significance level, in line with Hausman et al. (1992)[5]. Generally speaking, our estimates confirm most of the microstructure hypotheses described in Section 5.4.2. This is a first relevant result, which extends the previous empirical findings to the trading features of Euronext. In fact, as was anticipated in the Introduction, the prior research has disregarded some of the peculiarities of Euronext, such as the anonymous trading regime or the presence of market operators acting as liquidity providers (see Ranaldo, 2004 and Cao et al., 2009). Our principal results may be summarized as follows:

- *Trade-off between price and quantity.* The estimates provide evidence for a unidirectional effect between the two components of order aggressiveness. Most of q_t coefficients in Table 5.9 are not significant, while Table 5.10 always exhibits a positive and significant impact of Δp_t on the order size. Once controlled for the relevant LOB variables, our findings indicate that price aggressiveness is only marginally affected by quantity. Conversely, Lo and Sapp (2010) find a negative and significant impact of quantity on price aggressiveness. Although the direction of the price-volume decision is conflicting in the two papers, the inverse relationship between price and volume aggressiveness is definitely similar. Actually, the trade-off between

[5] Volume, volume at the best quotes, book imbalance, and signed cumulative volume are expressed in thousands of shares, in order to display homogeneous estimates in terms of magnitude.

Table 5.9: Price aggressiveness: parameter estimates and p-values.

	Alstom				Axa				Crédit Agricole			
	Buy		Sell		Buy		Sell		Buy		Sell	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Constant	4.81E-02	0.00	5.9E-02	0.00	8.68E-03	0.00	1.1E-02	0.00	1.08E-02	0.00	1.3E-02	0.00
q_t	-3.58E-03	0.06	-3.2E-02	0.00	-3.24E-03	0.00	2.5E-03	0.00	7.18E-04	0.16	9.2E-05	0.86
Δp_{t-1}	1.89E-01	0.00	2.7E-01	0.00	1.94E-01	0.00	8.9E-02	0.00	1.67E-01	0.00	1.3E-01	0.00
Δp_{t-2}	7.20E-02	0.00	1.3E-01	0.00	1.00E-01	0.00	6.2E-02	0.00	7.10E-02	0.00	8.3E-02	0.00
Δp_{t-3}	5.30E-02	0.00	1.1E-01	0.00	8.20E-02	0.00	4.0E-02	0.00	7.10E-02	0.00	6.3E-02	0.00
Δp_{t-4}	5.60E-02	0.00	1.1E-01	0.00	7.80E-02	0.00	3.6E-02	0.00	4.50E-02	0.00	5.8E-02	0.00
Same side	-8.43E-03	0.00	-1.1E-02	0.00	-3.17E-04	0.00	-4.3E-04	0.00	-2.72E-04	0.00	-2.6E-04	0.00
Opp. side	1.59E-03	0.00	-5.3E-04	0.38	2.06E-04	0.00	1.5E-04	0.00	1.22E-04	0.00	1.0E-04	0.00
Sigcum	-1.60E-02	0.00	1.1E-02	0.00	-1.32E-03	0.00	1.0E-03	0.00	-1.88E-03	0.00	1.8E-03	0.00
Spread	4.50E-01	0.00	4.6E-01	0.00	3.59E-01	0.00	5.1E-01	0.00	4.42E-01	0.00	4.4E-01	0.00
Volatility	1.23E+00	0.00	1.8E+00	0.00	3.37E+00	0.00	9.7E-01	0.00	1.60E+00	0.00	8.3E-01	0.00
Return	-1.97E+00	0.00	1.4E+00	0.00	-3.58E-01	0.00	-9.7E-02	0.02	-1.09E-01	0.01	-1.7E-01	0.00
Time	-2.00E-03	0.00	-3.0E-03	0.00	-5.77E-04	0.00	-2.2E-04	0.01	-6.00E-04	0.00	-2.3E-04	0.00
δ_t	-1.90E-02	0.00	-3.2E-02	0.00	-5.10E-03	0.00	-6.8E-03	0.00	-5.65E-03	0.00	-7.4E-03	0.00
D_{open}	1.81E-02	0.00	7.9E-03	0.01	5.59E-03	0.00	7.0E-03	0.00	4.32E-03	0.00	4.8E-03	0.00
D_{lunch}	-5.34E-04	0.51	-1.5E-04	0.89	1.10E-03	0.00	-1.1E-03	0.00	9.47E-04	0.00	-2.8E-04	0.39
D_{SP}	-2.30E-03	0.00	6.9E-03	0.00	-1.79E-04	0.37	7.6E-04	0.00	5.20E-04	0.01	6.2E-04	0.02
D_{clos}	3.94E-03	0.00	1.6E-02	0.00	1.44E-03	0.00	1.6E-03	0.00	5.14E-03	0.00	9.2E-03	0.00
Imbalance	1.11E-03	0.00	3.9E-03	0.00	2.83E-04	0.00	8.4E-05	0.00	-3.05E-05	0.18	9.6E-05	0.00

Table 5.9: Continued from the previous page.

	Eads				Essilor				Sanofi-Aventis			
	Buy		Sell		Buy		Sell		Buy		Sell	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Constant	1.44E-02	0.00	1.2E-02	0.00	7.03E-03	0.01	1.5E-02	0.00	2.17E-02	0.00	2.7E-02	0.00
q_t	4.47E-03	0.00	2.2E-03	0.02	1.57E-02	0.11	3.0E-02	0.00	4.71E-03	0.02	9.9E-04	0.51
Δp_{t-1}	1.17E-01	0.00	1.5E-01	0.00	2.31E-01	0.00	1.3E-01	0.00	1.45E-01	0.00	1.4E-01	0.00
Δp_{t-2}	4.90E-02	0.00	5.5E-02	0.00	1.09E-01	0.00	5.9E-02	0.00	4.70E-02	0.00	6.9E-02	0.00
Δp_{t-3}	2.60E-02	0.00	4.8E-02	0.00	7.20E-02	0.00	4.9E-02	0.00	4.80E-02	0.00	5.1E-02	0.00
Δp_{t-4}	1.80E-02	0.00	3.7E-02	0.00	4.60E-02	0.00	3.8E-02	0.00	4.30E-02	0.00	4.4E-02	0.00
Same side	-1.01E-03	0.00	-1.1E-03	0.00	-4.23E-03	0.00	-3.8E-03	0.00	-2.67E-03	0.00	-2.5E-03	0.00
Opp. side	3.20E-04	0.00	3.4E-04	0.00	2.02E-03	0.00	1.2E-03	0.00	1.92E-03	0.00	6.4E-04	0.00
Sigcum	-4.98E-03	0.00	4.6E-03	0.00	-1.05E-02	0.00	4.6E-03	0.01	-5.16E-03	0.00	2.1E-03	0.00
Spread	5.23E-01	0.00	5.5E-01	0.00	2.70E-01	0.00	4.0E-01	0.00	4.46E-01	0.00	5.4E-01	0.00
Volatility	8.51E-01	0.00	1.6E+00	0.00	1.63E+00	0.00	7.5E-01	0.00	1.30E+00	0.00	9.6E-01	0.00
Return	2.11E-01	0.00	1.5E-01	0.00	1.00E-01	0.71	-9.6E-01	0.00	-8.50E-02	0.45	6.0E-03	0.95
Time	-3.91E-04	0.00	-4.7E-04	0.00	-4.92E-04	0.00	-2.8E-04	0.00	-1.57E-03	0.00	-2.3E-03	0.00
δ_t	-5.65E-03	0.00	-6.4E-03	0.00	-5.87E-03	0.00	-1.3E-02	0.00	-6.45E-03	0.00	-5.4E-03	0.00
D_{open}	6.72E-03	0.00	4.8E-03	0.00	2.58E-02	0.00	1.9E-02	0.00	1.55E-02	0.00	1.4E-02	0.00
D_{lunch}	2.43E-04	0.33	-5.3E-04	0.15	5.14E-04	0.24	-8.4E-04	0.10	-9.94E-04	0.01	3.1E-03	0.00
D_{SP}	6.34E-04	0.02	1.2E-04	0.74	1.42E-03	0.00	8.2E-04	0.09	-5.03E-04	0.15	2.0E-04	0.49
D_{clos}	-1.24E-03	0.01	5.3E-04	0.47	-4.43E-03	0.01	-9.7E-05	0.96	1.29E-03	0.06	-1.4E-05	0.98
Imbalance	-5.28E-05	0.34	2.1E-04	0.00	2.48E-04	0.47	1.7E-03	0.00	6.22E-04	0.00	4.5E-04	0.00

quantity and price was already evident from Table 5.1 and it can be attributed to the order submission strategies. As it was previously discussed, the traders combine high price aggressiveness with small quantity to protect their private information and to ensure a fast and certain execution.

- *Bid-ask spread.* Table 5.9 uniformly exhibits a positive coefficient for the bid-ask spread, implying that a larger spread is associated with the submission of less aggressive orders. The result holds for both sides of the market and it is coherent with previous researches in this field (e.g. Ellul et al., 2003 or Ranaldo, 2004). The spread is perceived as a reliable measure of the presence of informational asymmetries. Therefore, a wider bid-ask spread induces the traders to submit less aggressive orders to minimize the risk of being picked off by agents with superior information. As to quantity, Table 5.10 mostly shows negative estimates, though no significant effect is found for Axa and Sanofi-Aventis. The estimates in Lo and Sapp (2010) also show that a wider spread is combined with a smaller order size. This confirms the assumption that traders are more likely to reduce their exposition in terms of quantity when the adverse selection risk is a reliable menace.

- *Same side depth.* Table 5.9 presents a negative and significant estimate for the variable that measures the depth on the own side of the market. This result is consistent with empirical contributions testing the crowding-out hypothesis (e.g. Beber and Caglio, 2005 or Duong et al., 2009), and it confirms the existence of a competition effect among traders. The result for quantity is more ambiguous, and it is non-significant to a large extent. Generally speaking, it is quite hard to identify a clear effect on the order size. The few significant estimates display a positive sign as in Lo and Sapp (2010); however, this does not match the negative impact that would be expected to speed up the order execution process.

- *Opposite side depth.* Coherent with the theoretical assumptions, the depth on the opposite side of the market is inversely related to price aggressiveness. The result is clearly shown in Table 5.9, which always exhibits a positive and significant coefficient for this regressor. These findings confirm the strategic effect described in Parlour (1998), and already encountered in Pascual and Veredas (2009) or Ranaldo (2004), among others. For the quantity, the percentage of significant estimates is slightly larger with respect to depth on the own side. In this case, the positive coefficients indicate that the traders submit larger orders to fully exploit the depth on the opposite side of the market.

- *Volatility.* We find evidence of a negative relationship between volatility and price aggressiveness, with Table 5.9 displaying a striking uniformity of results. All the coefficients are positive and significant, which is consistent with most of the previous empirical contributions as Bae et al. (2003) and Ahn et al. (2001); conversely, Cao et al. (2008) find a minimal effect for this variable. When market uncertainty is high, the traders submit less aggressive orders in terms of price, to minimize the risk of suffering a loss by trading with informed agents. The effect on quantity is also uniform across the six stocks, and Table 5.10 always presents positive and significant estimates, coherently with Ahn et al. (2001) and partially with Lo and Sapp (2010). These results are also in line with Handa and Schwartz (1996) who show that traders provide more liquidity when the transitory volatility is high. Actually, this would not match with the assumption of informed-based volatility that has been accounted for price aggressiveness. In any case, as anticipated in Section 5.4.2, the empirical literature has found alternative results for the relationship between volatility and quantity.

- *Return.* Our estimates are in line with Cao et al. (2008) who find a minimal effect of return over price aggressiveness, while they are contrary to Ellul et al. (2003) who find evidence of a momentum strategy. Table 5.9 exhibits four out of ten non-significant coefficients, and the remaining estimates do not display a uniform sign direction; nevertheless, the momentum effect seems slightly more prevalent.

Table 5.10: Volume aggressiveness: parameter estimates and p-values.

	Alstom				Axa				Crédit Agricole			
	Buy		Sell		Buy		Sell		Buy		Sell	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Constant	3.25E-01	0.00	4.44E-01	0.00	1.07E+00	0.00	6.33E-01	0.00	9.42E-01	0.00	7.25E-01	0.00
Δp_t	9.94E-01	0.00	2.09E-01	0.00	4.43E+00	0.00	8.60E+00	0.00	1.85E+01	0.00	1.43E+01	0.00
Q_{t-1}	8.90E-02	0.00	8.90E-02	0.00	6.90E-02	0.00	1.10E-01	0.00	4.90E-02	0.00	5.00E-02	0.00
Q_{t-2}	3.00E-02	0.00	3.30E-02	0.00	1.40E-02	0.00	2.00E-02	0.00	1.00E-02	0.00	2.20E-02	0.00
Q_{t-3}	1.90E-02	0.00	2.00E-02	0.00	8.00E-03	0.00	1.90E-02	0.00	9.00E-03	0.00	1.20E-02	0.00
Q_{t-4}	1.40E-02	0.00	7.00E-03	0.00	1.00E-02	0.00	1.70E-02	0.00	8.00E-03	0.00	1.40E-02	0.00
Same side	-1.20E-02	0.10	-1.70E-02	0.03	3.00E-03	0.24	5.00E-03	0.01	-1.00E-02	0.03	3.00E-03	0.40
Opp. side	3.00E-03	0.63	2.50E-02	0.01	1.00E-02	0.00	2.00E-03	0.25	-7.00E-03	0.07	-5.00E-03	0.12
Sigcum	-3.80E-02	0.00	3.70E-02	0.01	-6.00E-03	0.48	2.70E-02	0.01	-1.00E-03	0.98	5.10E-02	0.01
Spread	-1.15E+00	0.00	-6.58E-01	0.00	-7.21E-01	0.41	-3.23E-01	0.82	-9.94E+00	0.00	-6.96E+00	0.00
Volatility	1.58E+01	0.00	1.75E+01	0.00	5.34E+01	0.00	3.32E+01	0.00	1.08E+02	0.00	3.67E+01	0.00
Return	-6.72E-01	0.78	2.87E+00	0.21	-2.14E+01	0.00	-2.01E+01	0.00	3.29E+00	0.58	-1.62E+01	0.00
Time	-9.00E-03	0.00	-1.30E-02	0.00	-4.00E-03	0.03	2.00E-03	0.48	-2.50E-02	0.00	-7.00E-03	0.00
δ_t	-1.34E-01	0.00	-2.34E-01	0.00	-3.85E-01	0.00	-1.59E-01	0.00	-3.92E-01	0.00	-1.92E-01	0.00
D_{open}	1.92E-01	0.00	1.86E-01	0.00	-9.00E-02	0.01	-1.11E-01	0.00	1.39E-01	0.06	3.10E-02	0.53
D_{lunch}	3.00E-03	0.78	-4.00E-03	0.76	7.70E-02	0.00	2.30E-02	0.18	1.35E-01	0.00	8.00E-02	0.00
D_{SP}	-1.50E-02	0.06	-2.40E-02	0.00	3.00E-03	0.83	-2.10E-02	0.15	-1.20E-02	0.71	-8.00E-02	0.00
D_{clos}	5.00E-02	0.00	5.10E-02	0.00	1.24E-01	0.00	-2.40E-02	0.20	-3.70E-02	0.41	-1.71E-01	0.00
Imbalance	4.40E-02	0.00	5.20E-02	0.00	1.30E-02	0.00	1.00E-02	0.00	2.60E-02	0.00	1.30E-02	0.00

Table 5.10: Continued from the previous page.

	Eads				Essilor				Sanofi-Aventis			
	Buy		Sell		Buy		Sell		Buy		Sell	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
Constant	3.58E-01	0.00	5.50E-01	0.00	1.91E-01	0.00	1.65E-01	0.00	2.75E-01	0.00	3.21E-01	0.00
Δp_t	1.73E+01	0.00	1.18E+01	0.00	1.04E+00	0.00	1.44E+00	0.00	1.96E+00	0.00	9.39E-01	0.00
Q_{t-1}	8.30E-02	0.00	9.20E-02	0.00	1.21E-01	0.00	1.10E-01	0.00	8.20E-02	0.00	1.21E-01	0.00
Q_{t-2}	3.00E-02	0.00	2.20E-02	0.00	2.40E-02	0.00	3.70E-02	0.00	2.10E-02	0.00	3.10E-02	0.00
Q_{t-3}	8.00E-03	0.02	1.10E-02	0.00	2.70E-02	0.00	1.60E-02	0.00	1.20E-02	0.00	2.60E-02	0.00
Q_{t-4}	2.00E-03	0.54	8.00E-03	0.01	-3.00E-03	0.52	1.40E-02	0.01	1.10E-02	0.00	8.00E-03	0.00
Same side	-2.73E-03	0.71	-6.14E-03	0.25	9.36E-03	0.00	1.77E-03	0.85	6.24E-03	0.09	7.76E-03	0.00
Opp. side	1.52E-04	0.98	1.96E-02	0.00	1.28E-02	0.00	1.41E-02	0.00	1.05E-02	0.00	1.48E-03	0.68
Sigcum	-9.10E-02	0.01	-7.50E-02	0.02	-1.02E-01	0.00	5.00E-02	0.00	-1.30E-02	0.00	2.50E-02	0.00
Spread	-1.19E+01	0.00	-1.01E+01	0.00	-5.39E-01	0.00	-4.96E-01	0.01	2.78E-01	0.21	1.87E-01	0.33
Volatility	5.89E+01	0.00	7.06E+01	0.00	5.83E+00	0.00	6.12E+00	0.00	1.01E+01	0.00	1.16E+01	0.00
Return	2.99E+00	0.41	9.27E+00	0.03	3.75E+00	0.04	2.95E+00	0.14	-7.99E+00	0.00	4.36E+00	0.00
Time	-5.00E-03	0.00	-9.00E-03	0.00	0.00E+00	0.57	0.00E+00	0.06	4.00E-03	0.00	1.00E-03	0.37
δ_t	1.09E-01	0.01	3.50E-02	0.44	5.00E-02	0.00	4.80E-02	0.00	1.50E-02	0.11	-3.00E-02	0.00
D_{open}	2.70E-02	0.57	1.37E-01	0.00	3.40E-02	0.00	5.60E-02	0.00	-2.00E-03	0.83	1.10E-02	0.35
D_{lunch}	-4.20E-02	0.08	-2.90E-02	0.26	-2.20E-02	0.00	-1.00E-03	0.81	-4.00E-03	0.50	0.00E+00	0.99
D_{SP}	-1.18E-01	0.00	-8.10E-02	0.00	-1.60E-02	0.00	-1.50E-02	0.00	-8.00E-03	0.06	-1.10E-02	0.01
D_{clos}	1.00E-03	0.97	-1.40E-02	0.68	9.80E-02	0.00	6.90E-02	0.00	9.00E-02	0.00	6.20E-02	0.00
Imbalance	4.00E-02	0.00	2.80E-02	0.00	1.00E-03	0.79	1.30E-02	0.14	2.20E-02	0.00	2.60E-02	0.00

As to quantity, Table 5.10 displays a large majority of non-significant estimates; moreover, the few existing exceptions are not combined with significant estimates in terms of price aggressiveness. Generally speaking, it is fairly hard to single out an effect of return over the order size.

- *Time.* The results for time are partially in line with Ranaldo (2004). The traders usually associate a faster trading process with a high proportion of informed agents in the market; in this context, they are less likely to submit aggressive orders, since they are more exposed to the risk of being picked off by informed traders. Table 5.9 exhibits a negative coefficient for Δp_t in the majority of cases, though the effect is sometimes contrary to the predicted one. On the other hand, the impact on quantity is generally negative, as shown in Table 5.10. In general, when the trading frequency decreases and the presence of informed traders is presumably lower, we recover the usual combination of high price aggressiveness and small order size.

- *Temporal patterns.* The results reported in Table 5.9 validate the empirical findings for price aggressiveness discussed in Harris (1998) or BHS (1995). Our estimates strongly support the assumption that price aggressiveness increases throughout the day, with δ_k always displaying a negative and significant coefficient, see also Beber and Caglio (2005). This effect can be justified by the necessity, especially for institutional traders, of closing the open positions towards the end of the day. The estimate of δ_k is also mostly negative for volume aggressiveness, which confirms the tendency to combine a small order size with price aggressive orders to foster the execution process. On the other hand, the four time dummies exhibit a non-homogeneous outcome. Tables 5.9 and 5.10 show that D_{open} is usually associated with large limit orders, as in Bae et al. (2003) or Al-Suhaibani and Kryzanowski (2000). Conversely, it is not immediate to recover a clear effect for D_{lunch} and D_{SP}, both in terms of price and volume aggressiveness. D_{SP} generally display a negative effect on quantity, coherent with the fast execution process during the opening of the U.S. market; however, it is generally not combined with a higher price aggressiveness. Finally, and quite interestingly, D_{clos} mostly exhibits a positive coefficient for both Δp_t and volume aggressiveness. This result seems contradictory with respect to our findings for δ_k; however, a plausible explanation may be related to the high probability of informed-based trading in the last 30 minutes of the auction, which discourages the submission of price aggressive orders.

- *Lagged dependent variable.* Tables 5.9 and 5.10 display a striking homogeneity of results for the lagged dependent variables, with positive and strongly significant coefficients for all the lags. This result matches the previous findings on order autocorrelation (e.g. BHS, 1995, Pascual and Veredas, 2009) both in terms of price and volume.

- *Signed cumulative volume.* The signed cumulative volume shows an opposite effect for the two sides of the market. As expected, a positive pressure generates a negative effect on Δp_t for the bid side, and a positive impact for the ask side. This result is in line with the imitation effect and the momentum strategy documented in previous studies (e.g. BHS, 1995 or Ellul et al., 2003). Table 5.10 shows an analogous pattern for quantity: the estimates generally exhibit a negative sign for the buy side and a positive sign for the sell side. The traders combine their decisions in terms of price and volume, and submit more aggressive orders with smaller quantity to reduce the risk of non-execution.

- *Book imbalance.* In terms of price aggressiveness, Table 5.9 mostly shows positive estimates, which is consistent with a dominant strategic effect. Indeed, when the depth asymmetry increases, the traders are found to submit more limit orders, coherent with Parlour (1998). The result for quantity strictly follows, as a higher imbalance is generally associated with a larger order size, in line with the findings obtained for the opposite side depth. On the whole, both effects push to reduce the disequilibrium between the two sides of the market.

5.5.2 Order aggressiveness and price impact

As a final contribution of this research, we examine the price impact for different categories of orders. Our approach follows the one described in Beber and Caglio (2005) and Hopman (2007), among others. The price impact PI is computed as the absolute log return evaluated at the mid-quotes:

$$PI_{t+k} = |ln(m_{t+k}) - ln(m_t)|$$

where m_t is the mid-quote at time t, and k represents the time horizon. In our analysis, we separately compute the price impact for buy and sell orders, and we select four increasing time intervals k, corresponding to 1, 5, 15, and 30 minutes. The price impact is strictly related to the amount of information in the market, and it includes both a transient and a permanent component, see e.g. Hasbrouck and Seppi (1991). The former identifies the short-term effect related to temporary variations in the liquidity of the LOB; the latter spots a stable variation generated by the arrival of new information in the market. We separately compute the price impact for each category of order, on the basis of the Δp_t distribution intervals. To test the hypothesis that each category of order conveys a different amount of information, we assume that informed traders also adopt limit orders; this is in line with some previous findings (e.g. Anand et al., 2005), and it implies that not only trades may permanently affect the stock price. Moreover, we also check the assumption that most aggressive orders exert a greater price impact, coherent with the empirical findings in Beber and Caglio (2005) and Hopman (2007).

The results of our investigation are presented in Figures 5.2 and 5.3, which report the price impact as a non-parametric function of the order volume. We select Alstom (sell orders) and Eads (buy orders) as representative cases, though the remaining series exhibit similar patterns. For the sake of immediacy, we opt for a graphical representation of our findings, coherent with Potters and Bouchaud (2003) and Hopman (2007), among others. As a robustness check, we estimate a nonlinear regression of the price impact on

Figure 5.2: Price impact for Alstom buy orders. Each dot is a percentile of the volume distribution for a specific category of order. The green dots correspond to marketable limit orders, the red dots to market orders and the blue dots to limit orders. The curve is estimated as kernel regression with Epanechnikov kernel. The vertical axis displays the price variation expressed as the number of ticks, the horizontal axis provides the volume expressed as the number of shares.

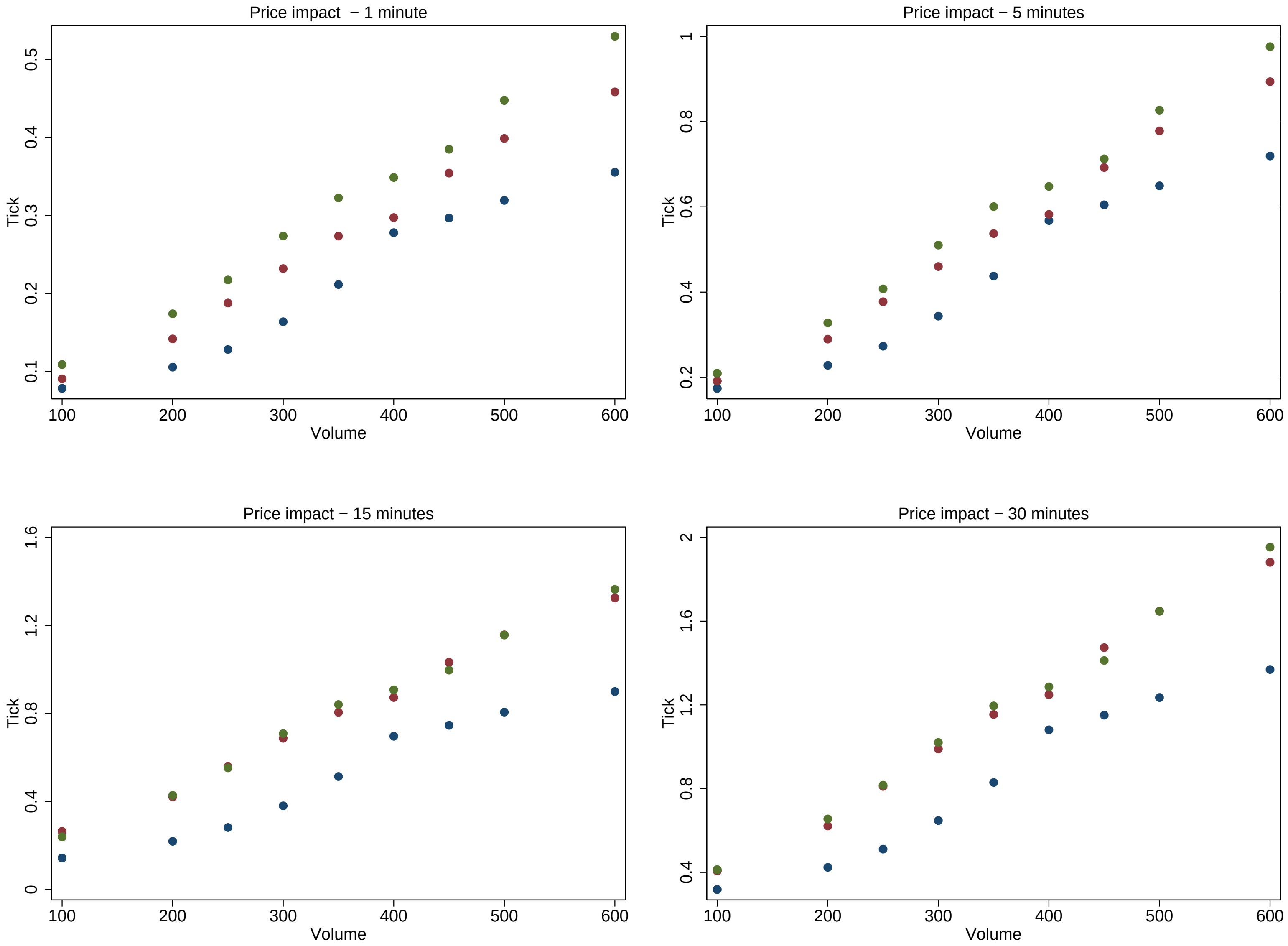
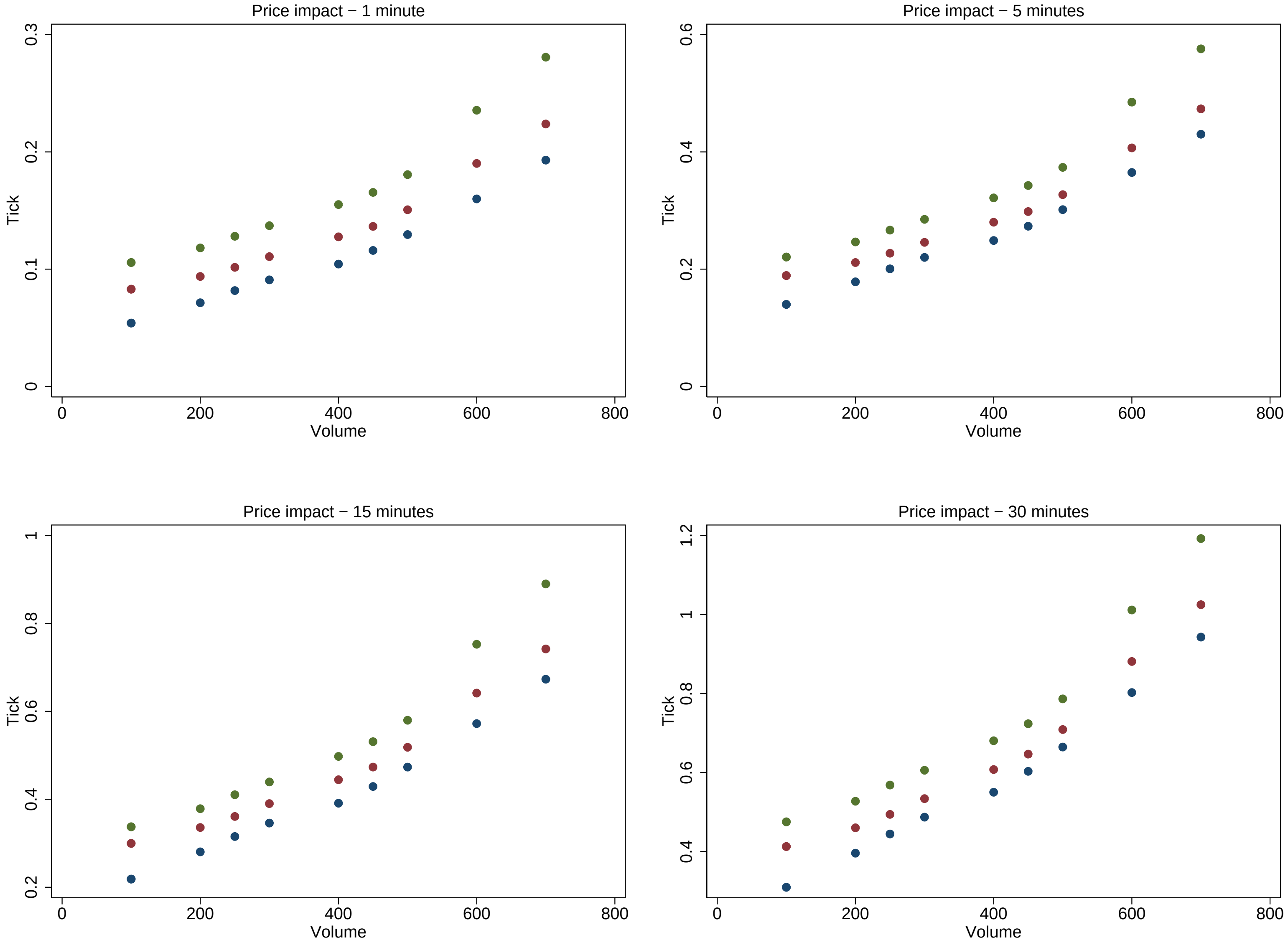


Figure 5.3: Price impact for Eads sell orders. Each dot is a percentile of the volume distribution for a specific category of order. The green dots correspond to marketable limit orders, the red dots to market orders and the blue dots to limit orders. The curve is estimated as kernel regression with Epanechnikov kernel. The vertical axis displays the price variation expressed as the number of ticks, the horizontal axis provides the volume expressed as the number of shares



order volume and dummies for Δp_t intervals, as in Hopman (2007). Our estimates show that marketable limit orders and market orders produce a larger pressure on prices with respect to limit orders; the results are available upon request. By observing the plots in Figures 5.2 and 5.3, we retrieve two main results. First, the higher the level of order aggressiveness, the larger the price impact; this holds constantly and independently of the time interval k. Beber and Caglio (2005) attribute this finding to the fact that marketable limit orders and market orders exercise a direct and immediate effect on the standing quotes and the depth of the LOB, while limit orders only exert an indirect pressure on the stock prices. Clearly, when price aggressive orders are only liquidity motivated, the price should revert to its previous value and the price impact should be rapidly absorbed by the market. Second, the price impact increases with larger quantity and longer time horizons. The direct relationship between the price impact and the order size is coherent with the previous empirical results and it is likely to be attributed to volume informativeness (see Easley and O'Hara, 1987). On the other hand, the relationship between price impact and time shows that there exists a (larger) long-term effect beyond the one in the short-run described by Beber and Caglio (2005). This long-term effect is mainly related to the informational content of the order flow, which is proven to be slowly absorbed by the market. Quite interestingly, we notice that the whole PI curve shifts upwards in the long term, i.e. when the price impact is more easily attributed to the presence of new information in the market. This is true even in the case of limit orders, which seems to further confirm that informed traders also adopt this type of orders. However, this last point, as well as the analysis of the price impact components with respect to order categories is left for future research.

5.6 Conclusions

This paper analyses the submission of orders at Euronext Paris, both in terms of price and volume. We propose a simple approach to express price aggressiveness in quantitative terms, which is made possible by the fact that price discreteness is a negligible issue in our data set. Our classification allows for a better representation of the order heterogeneity, since it quantitatively measures the price aggressiveness on the basis of the contingent market conditions. Although Euronext is ruled by strict price priority, this paper also examines the role of quantity, which represents a relevant indicator of order aggressiveness when the price informational content is limited. We adopt a simultaneous equations model to investigate the impact of a wide set of LOB variables on the order submission strategies. The empirical results confirm the theoretical assumptions of the literature, especially the presence of a strong autocorrelation pattern and the existence of daily cycles in both price and volume aggressiveness. Our estimates also confirm the inverse relationship between price aggressiveness and order size, with a one-way effect from price to quantity. We find price aggressiveness to be inversely related to the presence of informational asymmetries in the market, being mainly influenced by depth at best quotes, spread, volatility, and return. On the other hand, and as a probable consequence of the market trading rules, the volume plays a minor role in determining the overall level of aggressiveness. In fact, it is only marginally affected by LOB variables, with the main exception of spread and volatility. Finally, we also find the most aggressive orders to exert a higher impact on the stock prices.

Bibliography

Ahn, H.J., Bae, K.H., Chan, K. (2001), "Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong ," *The Journal of Finance*, 55, 769-790.

Aitken, M., Almeida, N., Harris F.H. deB., McInish, T.H. (2005), "Order Splitting and Order Aggressiveness in Electronic Trading," *Working paper.*

Aitken, M., Cook, R.M., Harris F.H. deB., McInish, T.H. (2007), "The Role of Market Makers in Electronic Markets: Liquidity Providers on Euronext Paris," *Working paper.*

Al-Suhaibani, M., Kryzanowski, L. (2000), "An Exploratory Analysis of the Order Book, and Order Flow and Execution on the Saudi Stock Market," *Journal of Banking and Finance*, 24, 1323-1357.

Anand, A., Chakravarty, S., Martell, T. (2005), "Empirical Evidence on the Evolution of Liquidity: Choice of Market versus Limit Orders by Informed and Uninformed Investors," *Journal of Financial Markets*, 8, 265-287.

Bae, K.H., Jang, H., Park, K.S. (2003), "Traders' Choice Between Limit and Market Orders: Evidence from NYSE Stocks," *Journal of Financial Markets*, 6, 517-538.

Beber, A., Caglio, C. (2005), "Order Submission Strategies and Information: Empirical Evidence from the NYSE", FAME Research Paper (146)

Biais, B., Hillion, P., and Spatt, C. (1995), "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *The Journal of Finance*, 50, 1655-1689.

Bisière C., Kamionka T. (2000), "Timing of Orders, Orders Aggressiveness and the Order Book at the Paris Bourse," *Annales d'Economie et de Statistique*, 60, 43-72.

Bloomfield, R., O'Hara M., Saar, G. (2005), "The Make or Take Decision in an Electronic Market: Evidence on the Evolution of Liquidity," *Journal of Financial Economics*, 75 165-199.

Cao, C., Hansch, O. and Wang, X. (2008), "Order Placement Strategies in a Pure Limit Order Book Market," *The Journal of Financial Research*, 31, 113-140.

Cao, C., Hansch, O. and Wang, X. (2009), "The Information Content of an Open Limit Order Book," *The Journal of Futures Markets*, 29 (1), 16-41.

Chakravarty, S. (2001), "Stealth Trading: which Traders' Trades Move Stock Prices?," *Journal of Financial Economics*, 61, 289-307.

Copeland, T. and D. Galai, (1983), "Information Effect on the Bid-Ask Spread," *The Journal of Finance*, 38, 1457-1469.

De Winne, R. and D'Hond, C. (2005), "Market Transparency and Traders' Behavior: an Analysis on Euronext with Full Order Book Data," *Working paper.*

Duong, H.N., Kalev, P.S., and Krishnamurti C. (2009), "Order Aggressiveness of Institutional and Individual Investors," *Pacific-Basin Finance Journal*, 17, 533-546.

Easley, D., O'Hara, M. (1987), "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19(1), 69-90.

Easley, D., O'Hara, M. (1992), "Time and the Process of Security Price Adjustment," *The Journal of Finance*, 47, 577-605.

Ellul, A., Holden C.W., Jain, P. and Jennings R. (2003), "Order Dynamics: Recent Evidence from the NYSE," *Journal of Empirical Finance*, 14, 636-661.

Ferriani, F. (2012), "Informed and Uninformed Traders at Work: Evidence from the French Market," *Munich Repec working paper*.

Foucault, T. (1999), "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market," *Journal of Financial Markets*, 2, 99-134.

Foucault, T., Kadan, O., Kandel, E. (2005), " Limit Order Book as a Market for Liquidity?," *The Review of Financial Studies*, 18(4), 1171-1217.

Foucault, T., Moinas, S., Theissen, E. (2007), " Does Anonymity Matter in Electronic Limit Order Markets?," *The Review of Financial Studies*, 5, 1707-1747.

Glosten, L.R., Milgrom, P.R. (1985), "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14(1), 71-100.

Goyenko, R.Y., Holden C.W., and Trzcinka, C.A. (2009), "Do Liquidity Measures Measure Liquidity ?," *Journal of Financial Economics*, 92, 153-181.

Griffiths, M.D., Smith B.F., Turnbull D.A.S., White R.W. (2000), "The Costs and Determinants of Order Aggressiveness," *Journal of Financial Economics*, 56, 65-88.

Hall, A.D., Hautsch, N. (2006), "Order Aggressiveness and Order Book Dynamics," *Empirical Economics*, 30, 973-1005.

Handa, P., Schwartz, R. (1996), "Limit Order Trading," *The Journal of Finance*, 51, 1835-1861.

Harris, L. (1998), "Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems,"*Financial Markets, Institutions and Instruments*, 7 , 1-75.

Harris, L., Hasbrouck, J. (1996), "Market vs. Limit Orders: the SuperDOT Evidence on Order Submission Strategy," *The Journal of Financial and Quantitative Analysis*, 31 , 213-231.

Hasbrouck, J. (2007), "Empirical Market Microstructure," *Oxford University Press*, 3, 1445-1477.

Hasbrouck, J. (2009), "Trading Costs and Returns for US Equities: Estimating Effective Costs from Daily Data," *The Journal of Finance*, 3, 1445-1477.

Hasbrouck, J., Saar, G. (2002), "Limit Orders and Volatility in a Hybrid Market: the Island ECN," *Working paper*.

Hasbrouck, J., Seppi, D. (2001), "Common Factors in Prices, Order Flows, and Liquidity," *Journal of Financial Economics*, 59, 383-411.

Hausman, J., Lo, A. and MacKinlay, A.C. (1992), "An Ordered Probit Analysis of Transaction Stock Prices," *Journal of Financial Economics*, 31, 319-379.

Hollifield, B., Miller, R.A., Sandas, P. (2004), "Empirical Analysis of Limit Order Markets," *Review of Economic Studies*, 71, 1027-1063.

Hopman, C. (2007), "Do Supply and Demand Drive Stock Prices?," *Quantitative Finance* 7, 37.

Lo, I. and Sapp, S.G. (2010), "Order Aggressiveness and Quantity: how are they Determined in a Limit Order Market?," *Journal of International Financial Markets, Institutions & Money*, 20, 213-237.

Parlour, C. (1998), "Price Dynamics in Limit Order Markets," *The Review of Financial Studies*, 11, 789-816.

Pascual, R., Veredas, D. (2009), "What Pieces of Limit Order Book Information Matter in Explaining Order Choice by Patient and Impatient Traders?," *Quantitative Finance* 9, 527-545.

Potters, M., Bouchaud, J.P. (2003), "More Statistical Properties of Order Books and Price Impact," *Physica A*, 324, 133-140.

Ranaldo, A. (2004), "Order Aggressiveness in Limit Order Book Markets," *Journal of Financial Markets*, 7, 53-74.

Veredas, D., Rodriguez-Poo, J., and Espasa, A. (2002), "On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: a Semiparametric Approach," *CORE Discussion Paper*, 23.

Xu, Y. (2009), "Order Aggressiveness on the ASX Market," *International Journal of Economics and Finance*, 1, 51-75.