

Intraday Transaction Price Dynamics

Serge DAROLLES, Christian GOURIÉROUX,
Gaëlle LE FOL*

ABSTRACT. – High frequency transaction prices exhibit two major characteristics: they are discrete in level and only exist at random transaction dates. In this paper, we seek to model transaction price dynamics, taking into account these two features. We specify the transaction price process as a Markov Chain with random transaction dates, and discuss various tools for dynamic analysis like the canonical decomposition, the scale and speed measures. The approach is applied to high frequency data on the stock Elf-Aquitaine traded on the Paris Bourse.

Dynamiques intra-journalières des prix de transaction

RÉSUMÉ. – Les prix de transaction intrajournaliers présentent deux caractéristiques majeures : ils sont discrets en niveau et n'existent qu'à des dates de transaction aléatoires. Nous proposons une modélisation de la dynamique des prix de transaction prenant en compte ces deux aspects. Nous modélisons le processus de prix à l'aide d'une chaîne de Markov et introduisons différents outils adaptés à l'étude de la dynamique, comme la décomposition canonique, les mesures d'échelle et de vitesse. Cette approche est utilisée pour étudier les cours de transaction de l'action Elf Aquitaine échangée à la Bourse de Paris.

* S. DAROLLES: Société Générale Asset Management, Hedge Funds Quantitative Research and CREST; C. GOURIÉROUX: CEPREMAP and CREST; G. LE FOL: EUREQUA, CNRS, Université de Paris I Panthéon-Sorbonne and CREST.

1 Introduction

The modeling of transaction price evolution must incorporate two stylized features about transaction price data. On the one hand, transaction dates are irregularly spaced and random. On the other hand, transaction prices take discrete values, which depend on the tick¹ used to quote the asset. However, the main part of the microstructure and high frequency literature consider these features separately.

The time factor has, for a long time, been ignored in the microstructure literature. In the (KYLE [1985]) model for example, orders are held up until a predetermined execution date and the orders arrival time is irrelevant. In the sequential model of (GLOSTEN, MILGROM [1985]), orders are not aggregated anymore, investors are randomly selected to trade in a population of potential traders. This random selection is independent of time. (DIAMOND, VERRECCHIA [1987]) are the first to consider time as a source of information and, in their model, no trade is bad news. Following (GLOSTEN, MILGROM [1985]), (EASLEY, O'HARA [1992]) propose a model such that the time factor is related to a random arrival of information on the market. This hypothesis allows them to incorporate time in the market-maker pricing rule and no trade means no news.

The same comments apply to the econometric high frequency literature. Many authors are interested in trading times and related liquidity analyses. It explains the studies on time deformation (see (MULLER *et al.* [1990]; MULLER *et al.* [1994]; DACOROGNA *et al.* [1994]), the introduction of activity and coactivity measures (GOURIÉROUX, JASIAK, LE FOL [1999]), and the development of duration models with lagged endogenous effects, either ACD models (ENGLE, RUSSEL [1994]; GHYSELS, JASIAK [1997]; ENGLE [2000]; BAUWENS, GIOT [1997]) or stochastic volatility duration models (GHYSELS, GOURIÉROUX, JASIAK [1997]). The analysis of price discreteness is a more recent topic and the modelling has to account for both admissible price values (HAUSMAN, LO, MACKINLAY [1992]) and investors preference for round numbers (see HARRIS [1990], DRAVID [1991]; HASBROUCK [1997]; MANRIQUE, SHEPHARD [1998]). The only papers to reconcile both features and propose a joint modelling of durations and discrete prices is the (ENGLE, RUSSEL [1998]) paper. Our paper is in the same spirit as their work.

In section 2, we analyze the intra-day price process through a joint modeling of intertrade durations and the associated trading prices. The model is compatible with the market efficiency assumption and is designed to capture the price discreteness. We then discuss the interpretation of several functional parameters such as the invariant measure, the scale function, the speed measure, the canonical decomposition, and explain how they can be used for analyzing mean reversion or asymmetries in the price movements. In section 3, we introduce the maximum likelihood estimator of the parameters

1. A tick is the minimum price increment allowed on financial markets.

of the price transitions and duration distributions, and describe their asymptotic properties. This approach is applied to a specific French stock, Elf-Aquitaine, traded on the Paris Bourse.

2 The Intraday Price Process

2.1 The Variables

We examine the intra-day transaction price process, *i.e.* the price levels and times within the day at which transactions take place. We omit the day index and we define by:

1. $n = 0, 1, \dots, N$, the transaction index, with $n = 0$ for the opening;
2. p_n the transaction price. If p_0 is the opening price, $X_n = p_n - p_0$ denotes the price variation from the market opening;
3. d_n the n^{th} transaction time, measured in seconds from the opening ($d_0 = 0$);
4. τ_n , $n = 1, \dots, N$, the durations between consecutive transactions ($n - 1$) and n .

In most microstructure models, market-makers update their quotes whenever they receive an order (see, for example, GLOSTEN, MILGROM [1985]; EASLEY, O'HARA [1987]; EASLEY, O'HARA [1992]; BOYER, LE FOL [1998]) prohibiting the occurrence of transactions with no price impact. Two consecutive trades occurring at the same price level imply that the market-makers did not change their quotes in the meanwhile and hence that the first received order was uninformative. Hence, it is also useful to consider a subsample of trading times corresponding to effective price movements. We define by:

1. $j = 0, 1, \dots, J$, the index of transactions with price movement;
2. N_j its index in the set of all transactions. The corresponding calendar dates are:

$$D_j = d_{N_j}.$$

3. The number of transactions with no price movement between two consecutive transactions with price movement, transactions $j - 1$ and j (say), is: $\Delta_j = N_j - N_{j-1} - 1$.
4. The durations between successive trades concluded at two distinct price levels are:

$$T_j = D_j - D_{j-1} = d_{N_j} - d_{N_{j-1}} = d_{\Delta_j+1} = \sum_{n=N_{j-1}+1}^{N_j} \tau_n.$$

5. The associated price levels adjusted for the opening are:

$$Y_j = X_{N_j}.$$

Briefly, our model involves three times, namely calendar time, transaction time and transaction time with price modifications.

2.2 The Intraday Price Process Distribution

Let us denote by $m, m = 1, \dots, M$, the days, and $i = 1, \dots, N_m$ observation indexes. We now investigate the distribution of:

$$N^m, p_0^m, p_1^m, \dots, p_{N_m}^m, d_1^m, \dots, d_{N_m}^m, m = 1, \dots, M,$$

or equivalently, the distribution of:

$$p_0^m, X_1^m, \dots, X_{N_m}^m, \tau_1^m, \dots, \tau_{N_m}^m, m = 1, \dots, M.$$

This set of variables involves both a daily opening price process (p_0^m, m varying), and intra-day trading processes (X_n^m, τ_n^m, n varying), $m = 1, \dots, M$. This separation is appealing for several reasons. First, in some financial markets, such as the New York Stock Exchange and the Paris Bourse for very active stocks, the opening price is determined by a call auction, whereas during the rest of the day the market works as a continuous auction system. These two systems are quite different (see GOURIÉROUX, LE FOL [1998]) for a description and a comparison of matching systems) as are the investors' behaviors. Thus, the trading organization implies that opening trades and intra-day trades should be modeled differently. Moreover, it is natural to separate the marginal distribution of the opening prices, which concerns a daily frequency, from the conditional distribution of the intra-day movements given the opening prices, which concerns the intra-day frequency.

Second, most of market microstructure models assume the existence of a daily fundamental value of the asset, say V^m . Let us consider a rational expectation approach. If we denote by I_0^m the public information about this value at the opening, then the opening price is $p_0^m = E(V^m | I_0^m)$. Note that I_0^m is an aggregate of past information about the daily fundamental value and hence is defined per day. Thereafter and all over the trading session, traders can infer about the fundamental asset value by watching the market. The price is then updated, $p_n^m = E(V^m | I_n^m)$, where

$$I_n^m = I_{n-1}^m + \{\text{information contained in trade } n\}.$$

In this rational expectation framework the updating, $E(V^m | I_n^m) - E(V^m | I_{n-1}^m)$ is not correlated with the initial price $E(V^m | I_0^m)$ (see SAMUELSON [1965]).

The assumptions below are defined in accordance with the previously described market organizations and microstructure models.

ASSUMPTION 2.1: The processes $(X_n^m, \tau_n^m, n$ varying), $m = 1, \dots, M$, are independent conditionally on the opening prices $p_0^m, m = 1, \dots, M$.

This assumption allows us to separate the analysis into two dynamics: the dynamic of opening prices and the dynamic of intra-day prices, and in parti-

cular to focus on the latter one by considering the distribution:

$$\prod_{m=1}^M l_m (X_n^m, \tau_n^m, n \text{ varying} \mid p_0^m, m = 1, \dots, M).$$

ASSUMPTION 2.2: Any intra-day process is independent of the opening price, and its distribution is independent of the considered day.

This assumption implies no daily seasonality for the process, or equivalently, the possibility to stationarize prices through the opening prices. The intra-day distribution becomes:

$$\prod_{m=1}^M l (X_n^m, \tau_n^m, n \text{ varying}) = \prod_{m=1}^M \prod_{n=1}^{N^m} l_n (X_n^m, \tau_n^m \mid \underline{X_{n-1}^m}, \underline{\tau_{n-1}^m}),$$

where $\underline{X_{n-1}^m}$ denotes the information contained in $X_0^m, X_1^m, \dots, X_{n-1}^m$, and the same notation applies to the durations.

ASSUMPTION 2.3: The market history influences the joint process exclusively through the last transaction price.

This assumption means that the last price contains all the relevant information by the market efficiency hypothesis (see FAMA [1970]). The efficiency hypothesis is assumed to hold during the day. Market inefficiency may occur due to the dynamics of the opening prices. Moreover, it is a basic assumption in all the continuous-time literature on complete markets, which are only defined in terms of prices (see section 4). In practice of course, this assumption should be tested in order to see whether past durations have an additional influence on price modifications (see EASLEY, O'HARA [1992]). The *informative price* assumption is symmetrical to the *informative duration* assumption considered in (ENGLE, RUSSEL [1998]), where lagged durations are assumed to contain all the relevant information. Under assumption 2.3, the intra-day distribution becomes:

$$\prod_{m=1}^M \prod_{n=1}^{N^m} l_n (X_n^m, \tau_n^m \mid X_{n-1}^m),$$

where X_0^m is set to 0, $m = 1, \dots, M$.

ASSUMPTION 2.4: The transaction process is homogeneous, *i.e.* the conditional distribution is independent of the transaction index.

This assumption implies no intraday seasonality on the price changes X_n^m . The conditional distribution becomes:

$$\prod_{m=1}^M \prod_{n=1}^{N^m} l (X_n^m, \tau_n^m \mid X_{n-1}^m),$$

and can be split up into:

$$\prod_{m=1}^M \prod_{n=1}^{N^m} l(X_n^m | X_{n-1}^m) l(\tau_n^m | X_{n-1}^m, X_n^m).$$

ASSUMPTION 2.5: Price modifications occur by ticks (set to 1), and the asset is liquid enough so that basic price movements cannot exceed one tick.

The first part of this assumption corresponds to an existing trading rule: prices on financial markets must be multiple of a price unit, called the tick. For the Paris Bourse in 1998, the tick size is given in table 1.²

Hence, any price movement is at least equal to the tick. The second part of the assumption states about the size of this price movement. The empirical analysis of intra-day price modifications shows that prices can definitely move by more than one tick. We give in Figure 1 the empirical distribution of intra-day price modifications, for Elf-Aquitaine, May, 1998. The price modifications range on that day, is $-5/+4$, with 91 % of the price modifications being $-1, 0$ or $+1$.

Three explanations may be given for the 9 % of larger jumps in prices. To understand the first one, let us describe briefly one characteristic of the French market. The Paris Bourse is an order driven market, *i.e.* without any intermediary ensuring liquidity and quoting buy and sell prices (bid and ask quotes). They are directly deduced from the order book by considering the standing limit orders. In this electronic system, any order that hits the market is either matched against a standing order or stored in the order book, by limit price and submission date, to be executed later on. Hence, the order book contains all standing buy and sell orders. The best supply and demand of the book are the standing sell order with the smallest price (ask price) and the standing buy order with the largest price (bid price). As an illustration, let us consider the arrival of a market order (*ordre à tout prix*) for a large volume. It will be executed against several standing orders clearing one side of the book, causing simultaneous trades at various prices and often forcing the average trading price to jump by more than one tick. In fact, these trades are not exactly simultaneous, but occur within a second. Because the second is the record step of the Paris Bourse, they are simultaneously recorded and, in econometric studies, these simultaneous trades are generally aggregated creating artificially large jumps in trading prices and zero intertrade durations. One data processing considers that these trades are equally spaced within a second and keep their various trading prices. For instance, five trades occurred at 10:46:45am, January the second 1998, are split up into five trades, the first one occurring at 10:46:45am, the second at 10:46:45 and 25 hundredth, the third at 10:46:45am and 50 hundredth, etc.

A second reason for large jumps in the transaction price is due to the spread. Typically the spread is the difference between the bid and ask prices,

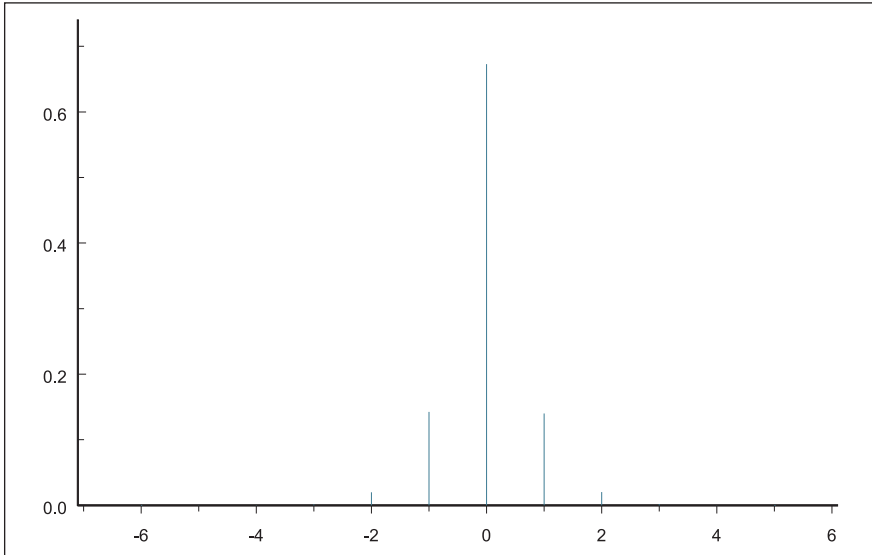
2. The tick size is now labelled in euro since January the 4th 1999 and set to: 0.01 euro for a stock price lower than 50 euro, 0.05 euro for a stock price between 50.05 and 100 euro, 0.10 euro for a stock price between 100.10 and 500 euro, 0.50 euro for a stock price greater than 500 euro.

TABLE 1

Paris Bourse Tick Size before January 1999

Stock Price	Tick
lower than 5 F	0.01 F
between 5.05 F and 100 F	0.05 F
between 100.10 F and 500 F	0.10 F
between 501 F and 5000 F	1 F
greater than 5000 F	10 F

FIGURE 1

Empirical Distribution of Price Modifications, May 1998, Elf-Aquitaine

i.e. the best standing demand and supply limiting prices. The transaction series (transaction price, traded volume, etc.) are collections of unsigned trades. By signed trades, we refer to buy or sell trades. If the spread is large and consecutive trades are not of the same type, the price modification can be larger than one tick. This phenomenon is the so-called bid and ask bounce. For instance, it happened between 10:07:41am and 10:10:23am, January the second 1998 for Elf-Aquitaine, where the transaction price evolution is given in Table 2 and pictured in Figure 2. The first trade is a sell, hits the bid price and clears the book on the buy side up to 700 FF, which becomes the new bid price. Hence, the spread becomes wider. The following trades are two buy orders at 703 FF. They do not modify the spread which remains 700-703 FF. Right after these buy orders, two sell orders arrive and are executed at the bid price (700 FF). This causes the transaction price to first move one tick up and then three ticks down, two ticks up, etc. One way to circumvent this problem due to the mixing of buy and sell transactions is to sign the trades. We call buy signed trade (or buy) any transaction initiated by a buyer against a sell standing order. Conversely, we call sell signed trade (or sell), any order

initiated by a seller and matched against a buy standing order. With this definition, an order executed at the bid price is a sell order and the transaction a sell, whereas an order executed at the ask is a buy order and the resulting trade a buy. From the high frequency transactions and the best limits data base, we construct a signed transactions data base, giving the day, the time, the price and whether the transaction was a buy or a sell (see Appendix C).

Finally, the signed transaction price can increase or decrease by more than one tick for liquidity or informational reasons. If the stock is liquid, there should be no empty spaces in the order book, and no empty price limit. Some price limits may be empty and consecutive buys or sells cause large jumps in the transaction price. It arises around 10:50am where the price fell from 700 FF to 698 FF. On the other hand, after a sequence of buy trades, the liquidity purveyors are updating their quotes upward, thinking that the event “*there is new information and it is good news*” is more likely (see EASLEY,

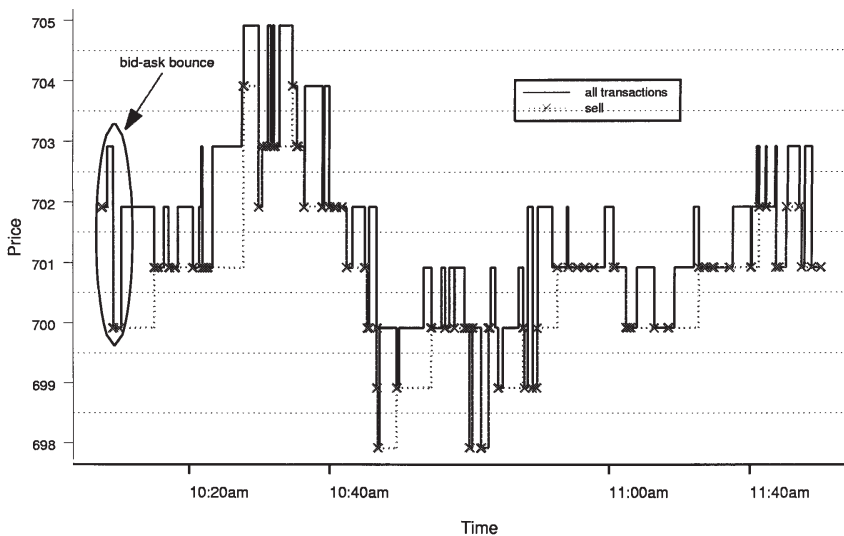
TABLE 2

Transaction Price Evolution, Elf Aquitaine, January the second, 1998

Time (hh:mm:ss)	Price (FF)	Spread bid-ask	Trade Type buy or sell
10:07:41am	702	702-703	sell
10:08:26am	703	700-703	buy
10:09:02am	703	700-703	buy
10:09:12am	700	700-703	sell
10:10:03am	700	700-703	sell
10:10:23am	702	701-702	buy

FIGURE 2

Transaction Price from 10:30am until 11:45am, January 2, 1998, Elf-Aquitaine



O'HARA [1987]; BOYER, LE FOL [1998]). This is what happened on Elf-Aquitaine around 10:25am where the transaction price between two sells has jumped from 701 FF to 704 FF. Such a large jump will be transformed into a sequence of one tick jumps and the information conveyed by one trade causing a two tick jump is assumed equivalent to the information conveyed by two one-tick-jump trades. The durations between these artificial one tick trades are obtained by dividing equally the intertrade duration. This data transformation is not too restrictive in our setting since we are interested in the time necessary to move the transaction price by one tick as well as the probability that this price modification is an increase or a decrease.

| ASSUMPTION 2.6: There exists a daily maximal price change, say \bar{x} ticks.

This assumption corresponds to an existing regulation rule. On financial markets, there exist upward and downward barriers applied with respect to the daily opening price. If the transaction price is moving up or down outside these barriers, the trading session is suspended.³ This implies that the $(X_n^m, n \text{ varying})$ process has a finite state space, with values:

$$-\bar{x}, -\bar{x} + 1, \dots, 0, \dots, \bar{x},$$

where we assume symmetric barriers. Moreover, since X_{n+1}^m can only take the three values $X_n^m - 1, X_n^m, X_n^m + 1$, the process $(X_n^m, n \text{ varying})$ is a birth-death Markov chain with a finite state space (see FELLER [1957]; KARLIN, TAYLOR [1975]). The correction of the intra-day price process by the opening price is essential to get daily invariant finite state space.⁴

The transitions are characterized by the transition matrix $P = [p(x, y)]$, where $p(x, y)$ denotes the probability to move from state x to state y . The P matrix can be written as:

$$\begin{bmatrix} p(-\bar{x}, -\bar{x}) & p(-\bar{x}, -\bar{x} + 1) & 0 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & p(x, x - 1) & p(x, x) & p(x, x + 1) & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & & & 0 & p(\bar{x}, \bar{x} - 1) & p(\bar{x}, \bar{x}) \end{bmatrix}.$$

The distribution of the $(X_n^m, n \text{ varying})$ process is then characterized by the quantities:

$$p(x, x - 1), p(x, x), p(x, x + 1), x \text{ varying}.$$

The conditional distributions of the durations τ_n^m , given the past and future states, must be specified to analyze the second part of the dynamics. Due to

3. On the Paris Bourse, for "continu A" securities, any time the transaction price is moving up by more than 21.25 % compared to the opening price, or moving down by more than 18.75 %, trading is suspended at least for 15 minutes.

4. Indeed, there is no reason to make such an assumption on the initial price process itself (see ENGLE, RUSSEL [1998] for such an assumption).

the data processing, we have no zero durations, and the distributions can be assumed to be continuous. We need three families of duration p.d.f. indexed by the previous state x :

$$g(\cdot | x, x-1), g(\cdot | x, x), g(\cdot | x, x+1), x \text{ varying.}$$

The intra-day dynamics can be rewritten to reveal the effect of the trades-with price movements. We get:

$$\begin{aligned} & \prod_{m=1}^M \prod_{j=1}^{J^m} \left[l\left(Y_j^m | Y_{j-1}^m\right) l\left(N_j^m | Y_{j-1}^m, Y_j^m\right) \prod_{n=N_{j-1}^m+1}^{N_j^m} l\left(\tau_n^m | X_{n-1}^m, X_n^m\right) \right] \\ &= \prod_{m=1}^M \prod_{j=1}^{J^m} \left[l\left(Y_j^m | Y_{j-1}^m\right) l\left(N_j^m | Y_{j-1}^m, Y_j^m\right) \prod_{n=N_{j-1}^m+1}^{N_j^m} l\left(\tau_n^m | Y_{j-1}^m, Y_j^m\right) \right]. \end{aligned}$$

The first term in the decomposition gives the distribution of the intradayprice process for the trades with price movements whereas the second one corresponds to the decomposition of spells without movement. From the classical properties of birth and death processes (see *e.g.* BREIMAN [1965], theorem 15.28, p. 329), we get:

PROPOSITION 2.1:

i) The $(Y_j^m, j \text{ varying})$ process is a birth and death process with transition probabilities:

$$q(x, x-1) = \frac{p(x, x-1)}{1-p(x, x)}, q(x, x+1) = \frac{p(x, x+1)}{1-p(x, x)},$$

ii) The conditional distribution of Δ_j given Y_{j-1}^m, Y_j^m depends only on the past through Y_{j-1}^m . If $Y_{j-1}^m = x$, this distribution is a geometric distribution with parameter $p(x, x)$.

Related distributions can easily be deduced. The distribution of the durations T_j^m between successive transactions concluded at two different price levels, given $Y_{j-1}^m = x, Y_j^m = x+1, \Delta_j$, is:

$$g(\cdot | x, x)^{* \Delta_j} * g(\cdot | x, x+1),$$

where $*$ refers to the convolution. The conditional distribution of T_j^m , given $Y_{j-1}^m = x, Y_j^m = x+1$ only, is deduced by integrating out Δ_j with respect to the geometric distribution with parameter $p(x, x)$; we get:

$$\sum_{n=0}^{\infty} \{ [1-p(x, x)] p(x, x)^n g(\cdot | x, x)^{*n} \} * g(\cdot | x, x+1).$$

Finally, we can consider the conditional distribution of Y_j^m given $T_j^m = T, Y_{j-1}^m$, defined by the transition probabilities $q(x, x+1; T)$ and

$q(x, x + 1; T)$, and study how they depend of T . If they are independent on T , no movement is no news. If $q(x, x + 1; T)$ is decreasing with T , no movement is bad news.

2.3 Parameters Interpretation

Before we tackle about the estimation problems or specify the model structure, it is useful to discuss parametrizations with financial interpretations. Several functional parameters will be introduced and linked to the invariant measure of the process, to the spectral decomposition of the transition matrix, to the scale function and to some speed measures.

2.3.1 Stationary Distribution and Equilibrium Reversion

The stationary distribution of the Markov chain $(X_n, n \text{ varying})$ is characterized by the proposition below, which is a version of the Chapman-Kolmogorov equation (see BREIMAN [1965], p. 319; and Appendix A for the proof).

PROPOSITION 2.2: The invariant measure $\pi(x)$, $-\bar{x} \leq x \leq \bar{x}$, and the transition probability $p(x, y)$, $-\bar{x} \leq x, y \leq \bar{x}$, satisfy the conditions:

$$\pi(x)p(x, y) = \pi(y)p(y, x),$$

for all $-\bar{x} + 1 \leq x, y \leq \bar{x}$. In particular, it is uniquely defined if: $p(x, x - 1) > 0$, for all x .

The previous relation is useful to analyze the property of reversion to a value x . Consider such a value, from which we can depart up (+ 1), or down (- 1). There is a local x -reversion if the probability to return to x is locally higher than to depart from it. This leads to the following definition:

DEFINITION 2.1: There is a local x -reversion if and only if:

$$p(x - 1, x) \geq p(x, x - 1) \text{ and } p(x + 1, x) \geq p(x, x + 1)$$

The proposition below is a direct consequence of proposition 2.2.

PROPOSITION 2.3: There is a local x -reversion, if and only if x is a mode of the stationary distribution.

In particular, we can consider cases where the stationary distribution admits a unique mode x^* . This condition is fulfilled if and only if: $p(x, x - 1) \geq p(x - 1, x)$, $\forall x < x^*$, and $p(x - 1, x) \geq p(x, x - 1)$, $\forall x > x^* + 1$. By analogy with the standard terminology,⁵ we say that there is a mode reversion.

5. Used for the Ornstein-Uhlenbeck process introduced for the dynamics of interest rates (see VASICEK [1977]).

The stationary distribution is informative on the relation between the opening and closing prices, p_0^m and p_c^m respectively. If the number of trades per day is large enough, the variable $p_c^m - p_0^m$ is independent on p_0^m , with approximative distribution the invariant distribution π . There is no *a priori* reason for this distribution to admit a zero mean (or mode), which would correspond to an intra-day random walk assumption. Therefore it is important to check if there exists at least one mode. In this case, it might be interesting to check if it coincides with the mean and if their common value is zero.

It is also possible to give other interpretations of multiple modes situations. They can prevail when there are investors with different behaviors and/or different informations in the market. These traders have various trading objectives and, as a consequence, the observed prices can switch from one type of price to another one depending on what type of trader is initiating the trade. Each mode can be interpreted as a potential traders price and the number of modes gives an idea of traders heterogeneity (see EPPS, EPPS [1976]). Several modes at 5 or 10 times the tick may also indicate investors preferences for round price values.

2.3.2 The Reversibility Property

The condition in proposition 2.2 admits an interesting interpretation. We have:

$$P [X_n = x \mid X_{n+1} = x - 1] = P [X_{n+1} = x \mid X_n = x - 1],$$

and:
$$P [X_n = x \mid X_{n+1} = x + 1] = P [X_{n+1} = x \mid X_n = x + 1],$$

which means that the (X_n , n varying) process is reversible (or symmetric), *i.e.* its dynamic properties are identical in the initial and reversed time. This property allows us for a simple analysis of the links between past and future as well as to alternative parametrizations of the chain by the so-called scale and speed functions (see the Feller decomposition, FELLER [1957]; SILVERSTEIN [1974]).

Note finally that the reversibility property can also be written as: $P'\Pi = \Pi P$, where P is the transition matrix and $\Pi = \text{diag } \pi$. In particular, the transition matrix admits real nonnegative eigenvalues.

2.3.3 Nonlinear Canonical Correlations

i) Canonical Correlation Analysis

Let us first recall the canonical correlation approach introduced by HOTELLING [1936]). We consider a pair of random vectors (X_1, X_2), both with dimension K (say). The variance-covariance matrix of (X_1', X_2') may be block-decomposed as:

$$(2.1) \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{21} = \Sigma_{12}'$. We are now interested in the linear combinations $a_1' X_1$ of components of X_1 and in the linear combinations $a_2' X_2$ of components of X_2 , which are the most correlated. Since the correlation is invariant when the

variables are multiplied by a scalar, we can normalize a_1 and a_2 to get unit variances for $a_1'X_1$ and $a_2'X_2$. These optimal combinations, called first canonical variates, are solutions of the constrained optimization problem:

$$(2.2) \quad \max a_1' \Sigma_{12} a_2,$$

subject to: $a_1' \Sigma_{11} a_1 = a_2' \Sigma_{22} a_2 = 1$. The first order conditions are derived by introducing Lagrange multipliers λ_1 and λ_2 associated with the constraints. The multipliers are identical at the optimum $\lambda_1 = \lambda_2 = \lambda$ (say), and the canonical directions are solutions of the system:

$$(2.3) \quad \begin{bmatrix} 0 & \Sigma_{12} \\ \Sigma_{21} & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \lambda \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix},$$

(see RAMSEY-SILVERMAN [1997]). When the matrices Σ_{11} and Σ_{22} are invertible, the formula (2.3) allows us for a separate computation of the vectors a_1 and a_2 . Let us introduce the canonical correlation matrices:

$$R^2(X_1, X_2) = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21},$$

and

$$R^2(X_2, X_1) = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12},$$

we get:

$$R^2(X_1, X_2) a_1 = \lambda^2 a_1,$$

and

$$R^2(X_2, X_1) a_2 = \lambda^2 a_2.$$

a_1 and a_2 are eigenvectors of the canonical correlation matrices associated with the largest eigenvalue to maximize the correlation. This procedure may continue to find the second canonical variates b_1, b_2 such that $b_1'X_1$ is not correlated with $a_1'X_1$, $b_2'X_2$ is not correlated with $a_2'X_2$, and $b_1'X_1$ and $b_2'X_2$ are the most correlated. It is known that the canonical variates of different orders are the eigenvectors of the correlation matrices in eigenvalues decreasing order.

ii) Application to Reversible Homogeneous Markov Chain

Let us now consider a reversible Markov chain (X_n, n varying) with a finite state space, $\{-\bar{x}, -\bar{x} + 1, \dots, 0, \dots, \bar{x}\}$ (say). We can associate to this chain the multivariate process (Z_n, n varying), whose dimension is $2\bar{x} + 1$ and components are the indicator functions of $X_n = x$, for $x = -\bar{x}, \dots, \bar{x}$. These components sum up to one. Next, we apply a canonical analysis to the pair of random vectors: $X_1 = Z_{n-1}, X_2 = Z_n$, first without centering the variables. The two canonical matrices are:

$$R^2(X_1, X_2) = E(Z_{n-1} Z'_{n-1})^{-1} E(Z_{n-1} Z'_n) E(Z_n Z'_n)^{-1} E(Z_n Z'_{n-1}),$$

and

$$R^2(X_2, X_1) = E(Z_n Z'_n)^{-1} E(Z_n Z'_{n-1}) E(Z_{n-1} Z'_{n-1})^{-1} E(Z_{n-1} Z'_n).$$

They are easily expressed in terms of the transition matrix P . Indeed, the regression of Z_n on Z_{n-1} is:

$$E[Z_n | Z_{n-1}] = E(Z_n Z'_{n-1}) E(Z_{n-1} Z'_{n-1})^{-1} Z_{n-1},$$

which implies:

$$P = E(Z_{n-1} Z'_{n-1})^{-1} E(Z_n Z'_{n-1}).$$

By the reversibility property, P is also equal to:

$$P = E(Z_n Z'_n)^{-1} E(Z_n Z'_{n-1}),$$

and we have:

$$R^2(X_1, X_2) = R^2(X_2, X_1) = P^2.$$

Thus, the canonical analysis is based on the spectral decomposition of the matrix P^2 , or equivalently P . Finally, the standard analysis with centered variables can be deduced from the analysis without centering the variables, as a consequence of the deterministic relation between the components of Z_n . Indeed, when performing the spectral decomposition of $R^2(X_1, X_2) = P^2$, the first eigenvalue is equal to one and the first eigenvector is proportional to the vector with unitary components. The next eigenvectors are orthogonal to the first one and therefore automatically centered. The canonical variates are obtained by retaining the eigenvalues and eigenvectors with an order larger than two.

2.3.4 Scale Function and Local Asymmetry of the Price Movement

It can also be of interest to study if the price movement is locally upward or downward. This can be achieved by comparing δ_x to 1, where:

$$(2.4) \quad \delta_x = \frac{p(x, x-1)}{p(x, x+1)} = \frac{q(x, x-1)}{q(x, x+1)}.$$

Prices are locally moving upward if $\delta_x < 1$, and conversely. The δ functions can be interpreted in terms of reaching times. Let us assume that the process is in state x at trade n , and denote by $n + N(a)$ the index of the first trading at which the process reaches the a -level. The probability for the process to escape downward is:

$$\gamma_x(x-1, x+1) = P^x [N(x-1) < N(x+1)] = \delta_x (1 + \delta_x)^{-1} = q(x, x-1).$$

We can generalize the result above and introduce the quantities:

$$(2.5) \quad \gamma_x(a, b) = P^x [N(a) < N(b)],$$

with $a < x < b, x, a, b \in \{-\bar{x}, \dots, +\bar{x}\}$. $\gamma_x(a, b)$ is the probability, starting from x , to escape downward from the (a, b) interval. Following BERGER [1993], theorem 5), we have the proposition below.

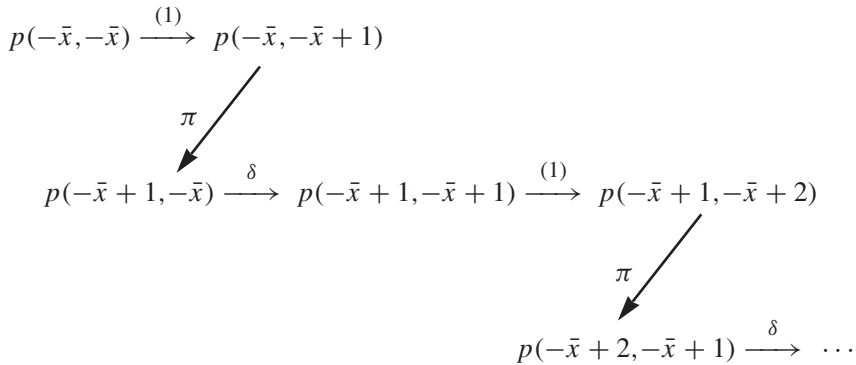
PROPOSITION 2.4:

$$\gamma_x(a,b) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y} = \frac{S(b) - S(x)}{S(b) - S(a)},$$

with $S(x) = \sum_{y=-\bar{x}+1}^{x-1} \gamma_y$, $\gamma_y = \delta_{-\bar{x}+1} \cdot \delta_{-\bar{x}+2} \dots \delta_y$, $x, a, b \in \{-\bar{x}, \dots, +\bar{x}\}$.

Remark 2.1

The functions π and δ can provide another parametrization of the transition matrix. To reconstruct the transition matrix, we just need π and δ , and for example the probability $p(-\bar{x}, -\bar{x})$, along the following scheme:



From the knowledge of $p(-\bar{x}, -\bar{x})$, we get $p(-\bar{x}, -\bar{x} + 1)$, since these two probabilities sum up to 1 (this condition is noted “(1)” in the scheme). Next, given this last probability and π we compute $p(-\bar{x} + 1, -\bar{x})$, and the knowledge of δ allows us to compute $p(-\bar{x} + 1, -\bar{x} + 1)$, etc.

2.3.5 Speed Measures

Various speed measures can be introduced depending on the selected time scale, either trading or calendar time. In transaction time, the speed measure is linked to the duration spent in a state x , which follows a geometric distribution with parameters $p(x, x)$. The expected duration $1/(1 - p(x, x))$ can be interpreted as an expected trading number to observe a price movement:

$$v_x(x - 1, x + 1) = E^x [N(x - 1) \wedge N(x + 1)] = \frac{1}{(1 - p(x, x))},$$

where $N(a)$ has been defined in the previous subsection, and:

$$(2.6) \quad v_x(a, b) = E^x [N(a) \wedge N(b)],$$

for $a < x < b$, $x, a, b \in \{-\bar{x}, \dots, +\bar{x}\}$. Following BERGER [1993], theorem 5), we get the proposition:

PROPOSITION 2.5:

$$v_x(a,b) = \frac{\sum_{y=a}^{x-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y} \sum_{y=a}^{b-1} \sum_{z=a}^y \frac{\gamma_y}{p_{z,z+1} \gamma_z} - \sum_{y=a}^{x-1} \sum_{z=a}^y \frac{\gamma_y}{p_{z,z+1} \gamma_z},$$

with $\gamma_y = \delta_{-\bar{x}+1} \cdot \delta_{-\bar{x}+2} \dots \delta_y$, $x, a, b \in \{-\bar{x}, \dots, +\bar{x}\}$.

A similar approach can be applied in calendar time. Let us introduce the reaching time $T(a)$ measured in calendar time. The expected waiting time for a price movement is: $v_x^c(x-1, x+1) = E^x [T(x-1) \wedge T(x+1)]$, and corresponds to a particular value of the function:

$$(2.7) \quad v_x^c(a,b) = E^x [T(a) \wedge T(b)], \quad x, a, b \in \{-\bar{x}, \dots, +\bar{x}\}$$

Remark 2.2

The speed measure in calendar time depends on the state x , on the transition probabilities, but also on the intertrade duration distributions. For instance, we have:

$$v_x^c(x-1, x+1) = \frac{1}{1 - p(x,x)} [p(x,x)\alpha(x,x) + p(x,x-1)\alpha(x,x-1) + p(x,x+1)\alpha(x,x+1)]$$

where $\alpha(x, x+j)$ is the mean of the duration distribution $g(\cdot; x, x+j)$.

2.3.6 The Duration Distributions

We can focus on some summary statistics of the intertrade duration distributions $g(\cdot | x, x-1)$, $g(\cdot | x, x)$, $g(\cdot | x, x+1)$, such as their means $\alpha(x, x-1)$, $\alpha(x, x)$, $\alpha(x, x+1)$ (if they exist). For example, we can study if the expected intertrade duration is larger when the price remains unchanged. This is achieved by comparing $\alpha(x, x+1)$ to $\alpha(x, x)$ and $\alpha(x+1, x+1)$.

3 Intraday Dynamic Estimation

In this section, we consider the maximum likelihood estimation of the intraday price distribution and apply these techniques to the high frequency data set corresponding to the Elf-Aquitaine stock.

3.1 Maximum Likelihood Estimator

The distribution of the Markov chain $(X_n, n \text{ varying})$ depends on a finite number of parameters. We assume that the conditional distribution of the durations τ_n^m given X_{n-1}^m, X_n^m belongs to a same parametric family $g(\cdot; \theta)$,

such that: $g(\cdot | x, x - 1) = g(\cdot; \theta(x, x - 1))$, $g(\cdot | x, x) = g(\cdot; \theta(x, x))$, $g(\cdot | x, x + 1) = g(\cdot; \theta(x, x + 1))$. The log-likelihood function can be written as:

$$\begin{aligned}
 L = & \sum_{m=1}^M \sum_{n=1}^{N_m} \left\{ \sum_{x=-\bar{x}}^{\bar{x}} I_{X_{n-1}^m=x, X_n^m=x} \log p(x, x) \right. \\
 & + I_{X_{n-1}^m=x, X_n^m=x+1} \log p(x, x+1) + I_{X_{n-1}^m=x, X_n^m=x-1} \log p(x, x-1) \\
 & + I_{X_{n-1}^m=x, X_n^m=x} \log g(\tau_n^m, \theta(x, x)) \\
 & + I_{X_{n-1}^m=x, X_n^m=x+1} \log g(\tau_n^m, \theta(x, x+1)) \\
 & \left. + I_{X_{n-1}^m=x, X_n^m=x-1} \log g(\tau_n^m, \theta(x, x-1)) \right\}
 \end{aligned}$$

where $I_{X_{n-1}^m=x, X_n^m=x}$ denotes the indicator function, which is equal to one if and only if $X_{n-1}^m = x$ and $X_n^m = x$. The log-likelihood has to be optimized under the constraint: $p(x, x) + p(x, x + 1) + p(x, x - 1) = 1$, $\forall x$. The proposition below results from the additive decomposition of the log-likelihood and the constraints with respect to the parameters.

PROPOSITION 3.6:

i) The maximum likelihood estimators below are asymptotically independent:

$$[\hat{p}(x, x - 1), \hat{p}(x, x), \hat{p}(x, x + 1)], \hat{\theta}(x, x - 1), \hat{\theta}(x, x), \hat{\theta}(x, x + 1),$$

x varying.

ii) Moreover, $\hat{p}(x, x)$ and $\hat{\delta}_x$ are asymptotically independent.

Thus we can consider independently the scale function parameters δ_x , x varying, and the parameters describing the calendar time speed of the price movements $p(x, x)$, $\theta(x, x - 1)$, $\theta(x, x)$, $\theta(x, x + 1)$. The estimators of the transition probabilities admit explicit forms and their asymptotic distributions are standard (see ANDERSON, GOODMAN [1957]; MADANSKY [1959]; BILLINGSLEY [1961]).

PROPOSITION 3.7:

i) The maximum likelihood estimators of the transition probabilities are equal to their empirical counterparts:

$$\hat{p}(x, x + j) = \frac{\sum_{m=1}^M \sum_{n=1}^{N_m} I_{X_{n-1}^m=x, X_n^m=x+j}}{\sum_{m=1}^M \sum_{n=1}^{N_m} I_{X_n^m=x}}, \quad j = -1, 0, 1,$$

$$\hat{\delta}_x = \frac{\hat{p}(x, x - 1)}{\hat{p}(x, x + 1)}.$$

ii) Let us denote by $N = \sum_{m=1}^M N_m$ the overall number of available observations on day m . We get the asymptotic distribution:

$$\sqrt{N} [\hat{p}(x, \cdot) - p(x, \cdot)] \xrightarrow{d} \mathcal{N} \left[0, \frac{1}{\pi(x)} [\text{diag } p(x, \cdot) - p(x, \cdot) p(x, \cdot)'] \right],$$

where:

$$p(x, \cdot) = \begin{bmatrix} p(x, x-1) \\ p(x, x) \\ p(x, x+1) \end{bmatrix}.$$

iii) In particular, we have:

$$\sqrt{N} [\hat{\delta}_x - \delta_x] \xrightarrow{d} \mathcal{N} \left[0, \frac{\delta_x^2}{\pi(x)} \left[\frac{1}{p(x, x-1)} + \frac{1}{p(x, x+1)} \right] \right].$$

A similar approach can be applied for the maximum likelihood estimators of the parameters $\theta(x, x+j)$, $j = -1, 0, 1$, involved in the intertrade durations distributions. These estimators are:

$$(3.1) \quad \hat{\theta}(x, x+j) = \arg \max_{\theta} \sum_{m=1}^M \sum_{n=1}^{N_m} I_{\{X_{n-1}^m=x, X_n^m=x+j\}} \log g(\tau_n^m, \theta),$$

ignoring the censored durations at the end of the day.

PROPOSITION 3.8: We have:

$$\sqrt{N} [\hat{\theta}(x, x+j) - \theta(x, x+j)] \xrightarrow{d} \mathcal{N} \left[0, \frac{1}{\pi(x) p(x, x+j)} I(\theta(x, x+j)) \right],$$

where $I(\theta)$ is the Fisher's information matrix associated with one observation and model $g(\cdot, \theta)$.

In simple cases, we get explicit forms of the estimators.

Example 3.1:

If the duration distribution is exponential:

$$g(y, \theta) = \theta \exp - \theta y,$$

we get:

$$\left[\hat{\theta}(x, x+j) \right]^{-1} = \frac{\sum_{m=1}^M \sum_{n=1}^{N_m} I_{\{X_{n-1}^m=x, X_n^m=x+j\}} \tau_n^m}{\sum_{m=1}^M \sum_{n=1}^{N_m} I_{\{X_{n-1}^m=x, X_n^m=x+j\}}}, \quad j = -1, 0, 1.$$

Example 3.2:

If the duration distribution follows a stable distribution, with parameter $\frac{1}{2}$:

$$g(y, \theta) = \frac{\theta^{\frac{1}{2}}}{\sqrt{2\pi y^3}} \exp\left(-\frac{\theta}{2y}\right) I_{y>0},$$

the estimator is given by:

$$\hat{\theta}(x, x+j) = \frac{\sum_{m=1}^M \sum_{n=1}^{N_m} \frac{1}{\tau_n^m} I_{\{X_{n-1}^m=x, X_n^m=x+j\}}}{\sum_{m=1}^M \sum_{n=1}^{N_m} I_{\{X_{n-1}^m=x, X_n^m=x+j\}}}.$$

The stable distributions are distributions of reaching times for a Brownian motion process and are natural candidates when comparing discrete and continuous time modelling. The mean of stable distributions does not exist, which explains why the maximum likelihood method is calibrated on the moment of $\frac{1}{\tau}$ instead of the τ one.

When the θ -parameters depend on the past through the initial state X_{n-1} only, *i.e.* $\theta(x, x-1) = \theta(x, x+1) = \theta(x, x)$, and the duration τ is exponentially distributed, durations can be aggregated and the distribution of the duration between consecutive trades, contracted at two distinct price levels, derived explicitly (see Appendix B).

PROPOSITION 3.9: If $\theta(x, x-1) = \theta(x, x+1) = \theta(x, x)$, and the intertrade durations are exponentially distributed with parameter $\theta(x, x)$, then the distribution of the durations T_j between consecutive trades, with effective price movement, is $\gamma(1, (1 - p(x, x))\theta(x, x))$.

3.2 Application

In this section, we estimate and discuss various functional parameters from the series of buy trades on Elf-Aquitaine, May 12, 1998. In order to get significant results, we aggregate the price extreme values and introduce the following bounds: -5 and 45 .

3.2.1 Transition Probabilities

We provide in Figures 3 and 4 the transition probabilities estimators as a function of the price level for the two processes (X_n^m , n varying) and (Y_n^m , n varying). We note the rather high probabilities to stay in the same state, and the increase of $p(x, x)$ with x . Under the martingale hypothesis, the probabilities $p(x, x-1)$ and $p(x, x+1)$ are equal for any x . Figure 3 shows that these two probabilities are close. However, the probabilities to exit upward are a bit larger than those to exit downward. This is confirmed by Figure 4, where

FIGURE 3

Transition Probabilities of X_n^m , for Elf-Aquitaine buy Trades

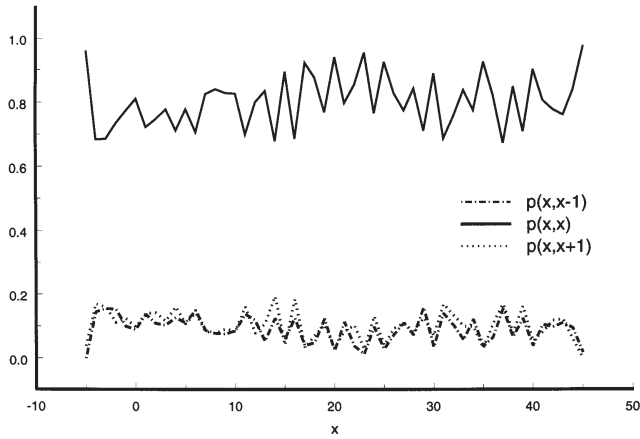


FIGURE 4

Transition Probabilities of Y_n^m , for Elf-Aquitaine buy Trades

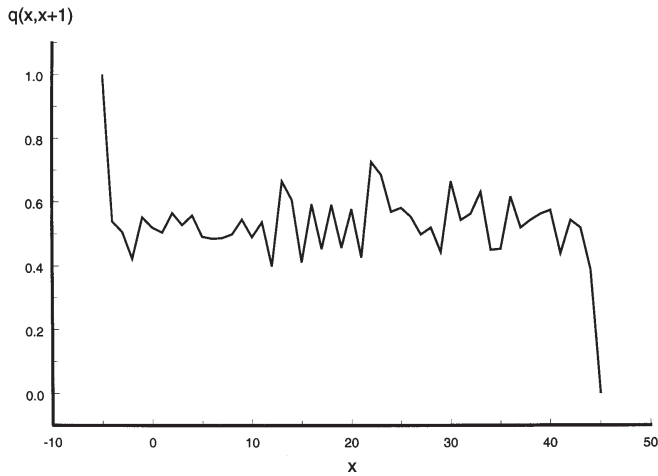


TABLE 3

Logit Approximation of the Transition Probabilities

Coefficients	Value	Standard error	t Value	$\Pr(> t)$
a_1	7.7674	2.0489	3.7910	0.0000
a_2	0.5636	0.2114	2.6658	0.0228
a_3	-0.0104	0.0054	-1.9286	0.1421
b_1	1.3778	0.10421	3.2204	0.0000
b_2	-0.0119	0.0108	-1.1056	0.2736
b_3	0.0001	0.0003	0.4750	0.6366

$q(x, x + 1)$ is on average larger than 0.5. The increase of the variability of the transition probabilities $q(x, x + 1)$ with x is likely an effect of the number of observations. When x is large, the number of movements starting from the price value x is rather small.

To structure these transition probabilities, we have estimated logit formulations including x and x^2 as explanatory variables:

$$p(x, x - 1) = \frac{1}{1 + e^{a_1 + a_2x + a_3x^2} + e^{b_1 + b_2x + b_3x^2}},$$

$$p(x, x) = e^{a_1 + a_2x + a_3x^2} p(x, x - 1),$$

$$p(x, x + 1) = e^{b_1 + b_2x + b_3x^2} p(x, x - 1),$$

and the q probabilities are such that:

$$q(x, x - 1) = \frac{1}{1 + e^{b_1 + b_2x + b_3x^2}}.$$

The estimation results are provided in Table 3.

We note that b_2 and b_3 are nonsignificant, and we find the monotonicity of $p(x, x)$ with respect to x due to the values of a_2 and a_3 .

3.2.2 The Invariant Measure

The invariant measure π is provided in Figure 5 without any aggregation of the extreme price levels. It shows that the smoothed distribution is unimodal, with a right skewness. This marginal skewness is due to the orientated bull market during this period, already observed in the transition probabilities.

3.2.3 Canonical Correlation Analysis

The first canonical correlations are close to one, equal to 1.00, 0.999, 0.998, 0.994, 0.990 respectively (see also Figure 6), whereas the corresponding canonical variates are plotted in Figure 7 as functions of the price level. They correspond to various transformations of the price, g_j , $j = 1, \dots, 4$, such that: $E[g_j(X_n) | X_{n-1}] \simeq g_j(X_{n-1})$. Therefore, these transformed series approximately satisfy the martingale condition.

The first non degenerate canonical direction V_2 is monotonous, close to the straight line. This corresponds to the usual intraday efficient market hypothesis, in which the prices themselves satisfy the martingale condition. The second non degenerate canonical direction corresponds to an asymmetric quadratic function. At this step, it is usual to find a transformation which measures the volatility effect. The high value of the associated eigenvalue reveals the volatility persistence, whereas the asymmetry corresponds to the so-called leverage effect.

3.2.4 The Intertrade Duration Distributions

The intertrade duration distributions can be determined for any initial price-value and any kind of price movement: up, stable or down. These distributions

FIGURE 5
Invariant measure A, for Elf-Aquitaine, May 1998

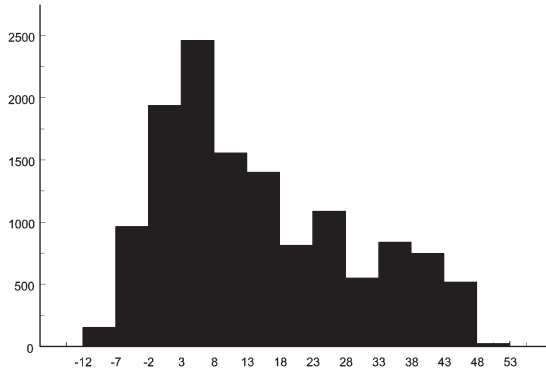


FIGURE 6
Eigenvalues of P^2 , for Elf-Aquitaine, May 1998

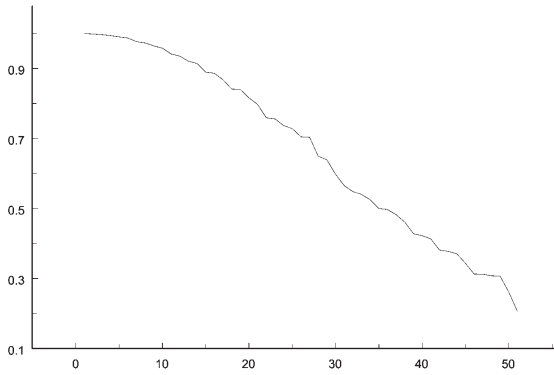
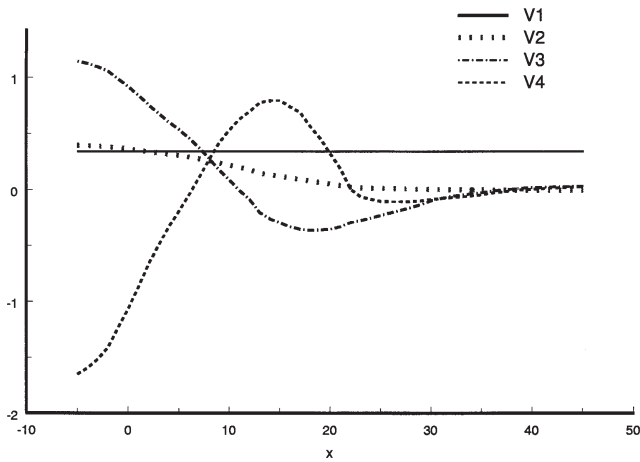


FIGURE 7
Eigenvectors of P^2 , for Elf-Aquitaine, May 1998



have been discretized by 10 seconds intervals and typical patterns are plotted in Figure 8 for the price level $x = 1$. The distributions corresponding to the “stable” movement are closer to exponential distributions, whereas the other ones often feature multimode.

These distributions vary with the price level x . To get some insight about these variations, we plotted in Figure 9 the means of the distributions: $\alpha(\cdot; x, x - 1)$, $\alpha(\cdot; x, x)$, $\alpha(\cdot; x, x + 1)$.

The comparison of these means (*i.e.* the expected times necessary to observe a new transaction one tick down and one tick up or a no price move-

FIGURE 8

Conditional distributions $g(\cdot; x, x - 1), g(\cdot; x, x), g(\cdot; x, x + 1)$, for $x = 1$, *Elf-Aquitaine, May 1998*

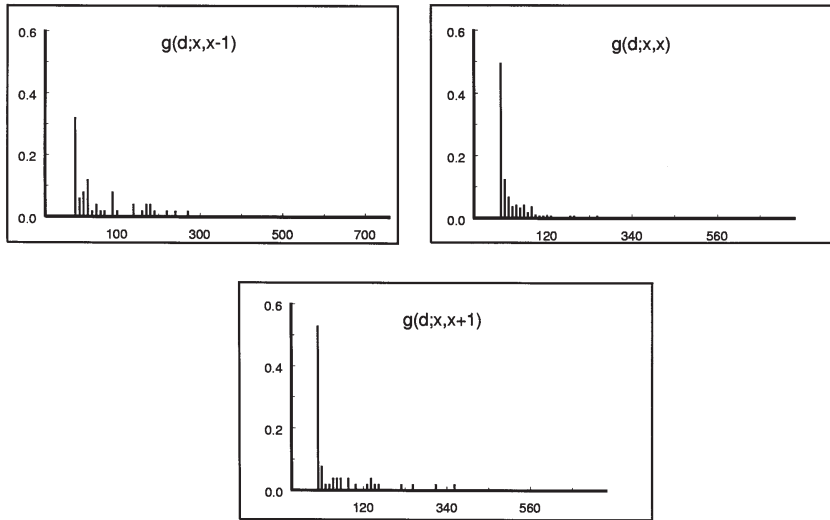
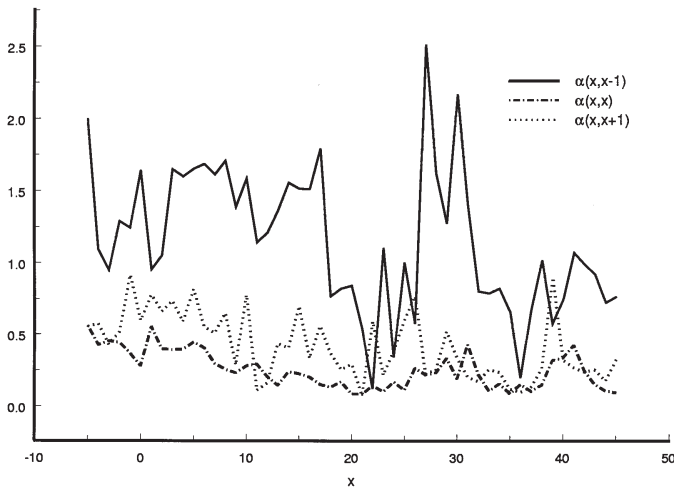


FIGURE 9

Means of duration distributions, Elf-Aquitaine, May 1998



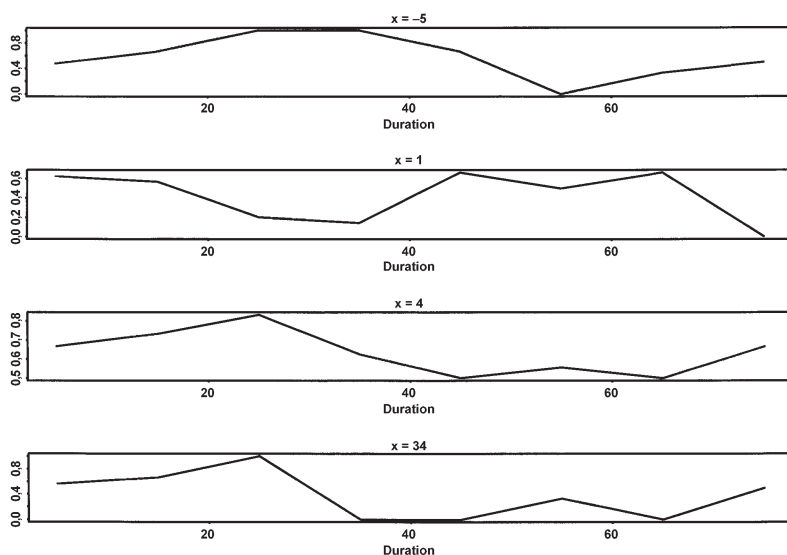
ment) has a direct interpretation in terms of news arrival. In the microstructure theoretical literature, some authors focussed on the role of time in the price adjustment. DIAMOND and VERRECCHIA [1987] conclude that, because of short sale constraints, an informed trader receiving a bad news will not trade when he would otherwise be selling and as a consequence a no trade period is a signal of bad news. A different explanation is given by EASLEY and O'HARA [1992]. In their model, the news arrival is an uncertain event and while trades provide a signal of either good or bad news, the absence of trade provides a signal of the existence of any new information event. In fact, if $\alpha(x, x - 1) < \alpha(x, x + 1)$, we can consider as DIAMOND and VERRECCHIA that *no trade is bad news*; if on average the duration to observe a decrease in the prices is slower than the average duration to observe an increase, then a no trade period (a long inter-trade duration period) should be followed by an increase in the prices (good news signal). Here, we can see that for Elf-Aquitaine, May 1998, the mean time to observe an up price movement or a no price movement is much lower than the one to observe a down price movement. Hence no trade is bad news.

3.2.5 Conditional Transition Probabilities

We also estimate the transition probabilities for effective price movements taking into account the duration. More precisely, we partition the set of values of the durations T_j^m , j varying, in consecutive intervals with the same probability 20%. We then compute the transition probabilities $q(x, x - 1; l)$, conditionally to the interval $l = 1, \dots, 5$, to which belongs the duration value. These transition probabilities are plotted in Figure 10 as functions of T_j^m for different price levels x . It is difficult to exhibit a systematic pattern, and to give any interpretation of a long stable movement as revealing bad news for instance.

FIGURE 10

Conditional Transition Probabilities, Elf-Aquitaine, May 1998



4 Concluding Remarks

We have introduced a joint specification for observed transaction prices and intertrade durations, which is compatible with the irregular spacing of the trades and the price discreteness. This joint distribution can be analyzed through several summary statistics, such as the scale and speed functions, or the spectral decomposition of the transition matrix. ■

• References

- ANDERSON T., GOODMAN L. (1957). – « Statistical Inference About Markov Chains », *Annals of Mathematical Statistics*, 28, pp. 89-109.
- BAUWENS L., GIOT P. (1997). – « The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process of Three NYSE Stocks », *Revised version of CORE DP 9789*. UCL, Louvain-La-Neuve.
- BERGER M. (1993). – *An Introduction to Probability and Stochastic Processes*, Springer, New-York.
- BILLINGSLEY P. (1961). – *Statistical Inference for Markov Processes*, Chicago University Press.
- BOYER C., LE FOL G. (1998). – « Random Time and Order Book Dynamics », *Cahier de la MSE*, 5, Paris 1 University.
- BREIMAN L. (1965). – *Probability*, Addison-Wesley.
- DACOROGNA M., GAUVREAU N.C., MULLER U.A., OLSEN R., PICTET O.V. (1994). – « Changing Time Scale for Short Term Forecasting in Financial Markets », Olsen and Associates, *Working Paper*, Zurich.
- DIAMOND D.W., VERRECCHIA R.E. (1987). – « Constraints on Short-selling and Asset Price Adjustments to Private Information », *Journal of Financial Economics*, 18, pp. 277-311.
- DRAVID A.R. (1991). – « Effects of Bid-Ask Spread and Price Discreteness on Stock Returns », *Working Paper*, Wharton School.
- EASLEY D., O'HARA M. (1987). – « Price, Trade Size, and Information in Securities Markets », *Journal of Financial Economics*, 19, pp. 69-90.
- EASLEY D., O'HARA M. (1992). – « Time and the Process of Security Price Adjustment », *Journal of Finance*, 47 (2), pp. 577-605.
- ENGLER R. (2000). – « The Econometrics of Ultra-High Frequency Data », *Econometrica*, 68 (1), pp. 1-22.
- ENGLER R., RUSSEL J., (1998). – « Autoregressive Conditional Duration; A New Model for Irregularly Spaced Times Series Data », *Econometrica*, 66 (5), pp. 1127-1162.
- ENGLER R., RUSSEL J., (1998). – « Econometric Analysis of Discrete-valued, Irregularly Spaced Financial Transaction Data using a New Autoregressive Conditional Multinomial Model », *Working Paper*, San Diego University.
- EPPS T., EPPS M. (1976). – « The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture of Distribution Hypothesis », *Econometrica*, 44, 2, pp. 305-325.
- FAMA E. (1970). – « Efficient Capital Markets: A Review of Theory and Empirical Work », *Journal of Finance*, 25, pp. 383-417.
- FELLER W. (1957). – *An Introduction to Probability Theory and Its Applications*, 2nd edition, Wiley, New-York.
- GHYSELS E., GOURIÉROUX C., JASIAK J. (1997). – « Stochastic Volatility Duration Models », *Working Paper*, CREST 9746.

- GHYSELS E., JASIAK J. (1997). – « GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model », *Working Paper 97s-06*, CIRANO, Montréal.
- GLOSTEN L., MILGROM P. (1985). – « Bid-Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders », *Journal of Financial Economics*, 14, pp. 71-100.
- GOURIÉROUX C., JASIAK J. (1998). – « Nonlinear Autocorrelograms with Application to Intertrade Durations », forthcoming *Journal of Time Series Analysis*.
- GOURIÉROUX C., JASIAK J. (2000). – *Financial Econometrics*, chap. 14, forthcoming, Princeton Univ. Press.
- GOURIÉROUX C., JASIAK J., LE FOL G. (1999). – « Intraday Market Activity », *Journal of Financial Markets*, 2, pp. 193-226.
- GOURIÉROUX C., LE FOL G. (1998). – « Matching Procedures and Market Characteristics », *Working Paper*, CREST 9815. French version in *Revue Économique*, 49, 3, pp. 795-808.
- HARRIS L. E. (1990). – « Estimation of Stock Variances and Serial Covariances from Discrete Observations », *Journal of Financial and Quantitative Analysis*, 25, pp. 291-306.
- HASBROUCK J. (1997). – « The Dynamics of Discrete Bid and Ask Quotes », *Working Paper*, Stern School of Business.
- HAUSMAN J., LO A., MACKINLAY C. (1992). – « An Ordered Probit Analysis of Transaction Stock Prices », *Journal of Financial Economics*, 31, pp. 319-330.
- HOTELLING H. (1936). – « Relations Between Two Sets of Variables », *Biometrika*, 28, pp. 321-377.
- KARLIN S., TAYLOR M. (1975). – *A First Course in Stochastic Processes*, Academic Press, London.
- KARLIN S., TAYLOR M. (1981). – *A Second Course in Stochastic Processes*, Academic Press, London.
- KYLE A. S. (1985). – « Continuous Auctions and Insider Trading », *Econometrica*, 53, pp. 1315-1336.
- LEE T., JUDGE G., ZELLNER A. (1968). – « Maximum Likelihood and Bayesian Estimation of Transition Probabilities », *JASA*, 63, pp. 1162-1179.
- MADANSKY A. (1959). – « Least Squares Estimation in Finite Markov Processes », *Psychometrika*, 24, pp. 137-144.
- MANRIQUE A., SHEPHARD N. (1998). – « Likelihood Analysis of a Discrete Bid/Ask Price Model for Common Stock Quoted on the NYSE », *Working Paper*, Nuffield College, Oxford University.
- MULLER U.A., DACOROGNA M.M., OLSEN R.B., PICTET O.V., SCHWARZ M., MORGENEGG C. (1990). – « Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis », *Journal of Banking and Finance*, 14, pp. 1189-1208.
- MULLER U.A., DACOROGNA M.M., DAVE R.D., OLSEN R.B., PICTET O.V., WARD J. (1994). – *Fractals and Intrinsic Time: A Challenge to Econometricians*, Olsen and Associates, DP, Zurich.
- O'HARA M. (1995). – *Market Microstructure Theory*, Blackwell.
- RAMSEY J.O., SILVERMAN B.W. (1997). – *Functional Data Analysis*, Springer Verlag, New York.
- SAMUELSON P. (1965). – « Proof that Properly Anticipated Prices Fluctuate Randomly », *Industrial Management Review*, 6, pp. 41-49.
- SILVERSTEIN M. (1974). – « Symmetric Markov Chain », *The Annals of Probability*, 2, pp. 681-701.
- VASICEK O. (1977). – « An Equilibrium Characterization of the Term Structure », *Journal of Financial Economics*, 5, pp. 177-88.

APPENDIX

A Invariant Measure Characterization

Using the Chapman-Kolmogorov equation for $x = -\bar{x}$, we get:

$$\begin{aligned}\pi(-\bar{x}) &= \pi(-\bar{x})p(-\bar{x}, -\bar{x}) + \pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x}) \\ &= \pi(-\bar{x}) [1 - p(-\bar{x}, -\bar{x} + 1)] + \pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x}),\end{aligned}$$

and

$$\pi(-\bar{x})p(-\bar{x}, -\bar{x} + 1) = \pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x}),$$

which corresponds to the expected relation for $x = -\bar{x}$ and $y = -\bar{x} + 1$. Moreover, we have:

$$\pi(-\bar{x})p(-\bar{x}, y) = \pi(y)p(y, -\bar{x}) = 0,$$

for any $y \geq -\bar{x} + 2$. Hence, we obtain the relation for $x = -\bar{x}$ and any y . For $x = -\bar{x} + 1$, we get:

$$\begin{aligned}\pi(-\bar{x} + 1) &= \pi(-\bar{x})p(-\bar{x}, -\bar{x} + 1) + \pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x} + 1) \\ &\quad + \pi(-\bar{x} + 2)p(-\bar{x} + 2, -\bar{x} + 1) \\ &= \pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x}) + \pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x} + 1) \\ &\quad + \pi(-\bar{x} + 2)p(-\bar{x} + 2, -\bar{x} + 1) \\ &= \pi(-\bar{x} + 1) [1 - p(-\bar{x} + 1, -\bar{x} + 2)] \\ &\quad + \pi(-\bar{x} + 2)p(-\bar{x} + 2, -\bar{x} + 1),\end{aligned}$$

and we obtain:

$$\pi(-\bar{x} + 1)p(-\bar{x} + 1, -\bar{x} + 2) = \pi(-\bar{x} + 2)p(-\bar{x} + 2, -\bar{x} + 1).$$

Since $\pi(-\bar{x} + 1)p(-\bar{x} + 1, y) = \pi(y)p(-\bar{x} + 2, -\bar{x} + 1) = 0$ for any $y \geq -\bar{x} + 3$, we have the relation $x = -\bar{x} + 1$ and any y . Now, if we suppose the relation is verified for $x - 1$, we obtain at rank x the relation:

$$\begin{aligned}\pi(x) &= \pi(x - 1)p(x - 1, x) + \pi(x)p(x, x) + \pi(x + 1)p(x + 1, x) \\ &= \pi(-x)p(x - 1, x) + \pi(x)p(x, x) + \pi(x + 1)p(x + 1, x),\end{aligned}$$

and

$$\pi(x)p(x, x + 1) = \pi(x + 1)p(x + 1, x).$$

A similar computation gives:

$$\pi(x)p(x, x - 1) = \pi(x - 1)p(x - 1, x).$$

Hence, we get the relation for any y .

B Distribution of the Duration between Price Modifications

If $\theta(x, x + 1) = \theta(x, x)$, and the duration distribution is exponential with $\theta(x, x)$, then:

$$\tau \sim \theta(x, x) \exp[-\theta(x, x)\tau].$$

Therefore the distribution of the duration between price modifications is:

$$\begin{aligned} & \sum_{n=1}^{\infty} (1 - p(x, x)) p(x, x)^{n-1} \frac{1}{(n-1)!} \theta(x, x)^n \exp[-\theta(x, x)\tau] \tau^{n-1} \\ &= (1 - p(x, x)) \theta(x, x) \sum_{n=1}^{\infty} \frac{p(x, x)^{n-1} \theta(x, x)^{n-1} \tau^{n-1}}{(n-1)!} \exp[-\theta(x, x)\tau] \\ &= (1 - p(x, x)) \theta(x, x) \exp[p(x, x)\theta(x, x) - \theta(x, x)\tau], \end{aligned}$$

which is the p.d.f. of a $\gamma(1, (1 - p(x, x))\theta(x, x))$ distribution.

C Signed Transactions

We describe in the tables below the available data, and the data transformations which have been performed. The first two tables provide informations on transactions: time, price, quantity, contract index [Table 1] and on the best limits of the order files [Table 2]. These data are used to sign the transaction, buy or sell, [Table 3] and then to select signed trades [Table 4].

TABLE 1
BDM, Elf Aquitaine

Transaction Data				
Date	Time	Seq. NB	Price	Quantity
05/01/1998	114436	1	714	1405
05/01/1998	114436	2	715	195
05/01/1998	114500	1	715	238
05/01/1998	114500	2	715	62
05/01/1998	114629	1	713	32
05/01/1998	114644	1	713	25
05/01/1998	114658	1	713	17
05/01/1998	114658	2	713	500
05/01/1998	114658	3	713	483
05/01/1998	114734	1	713	517
05/01/1998	114734	2	713	483
05/01/1998	114752	1	713	17
05/01/1998	114802	1	712	1
05/01/1998	114807	1	712	12
05/01/1998	114810	1	713	3
05/01/1998	114810	2	714	687
05/01/1998	114908	1	712	29
05/01/1998	114914	1	712	29
05/01/1998	114916	1	714	313
05/01/1998	114916	2	715	187
05/01/1998	114955	1	714	4740
05/01/1998	115045	1	713	200
05/01/1998	115143	1	714	400
05/01/1998	115218	1	714	100
05/01/1998	115324	1	714	4700
05/01/1998	115349	1	714	16
05/01/1998	115443	1	714	19
05/01/1998	115507	1	713	50

TABLE 2
BDM, Elf Aquitaine

Best Limit Data								
Date	Time	Nb Seq.	Bid	Q^b	Nb Sellers	Ask	Q^a	Nb Buyers
05/01/1998	114435	1	712	5180	6	714	1405	1
05/01/1998	114436	1	712	5180	6	715	5523	8
05/01/1998	114437	1	712	4680	5	715	5523	8
05/01/1998	114500	1	712	4680	5	715	5223	7
05/01/1998	114509	1	713	74	1	715	5223	7
05/01/1998	114519	1	713	574	2	715	5223	7
05/01/1998	114532	1	713	1574	3	715	5223	7
05/01/1998	114620	1	713	1574	3	715	6223	8
05/01/1998	114629	1	713	1542	3	715	6223	8
05/01/1998	114644	1	713	1517	3	715	6223	8
05/01/1998	114658	1	713	517	1	715	6223	8
05/01/1998	114712	1	713	1017	2	715	6223	8
05/01/1998	114734	1	713	17	1	715	6223	8
05/01/1998	114737	1	713	17	1	714	1000	1
05/01/1998	114752	1	712	5680	6	713	3	1
05/01/1998	114802	1	712	5679	6	713	3	1
05/01/1998	114807	1	712	5667	6	713	3	1
05/01/1998	114810	1	712	5667	6	714	313	1
05/01/1998	114908	1	712	5638	6	714	313	1
05/01/1998	114914	1	712	5609	6	714	313	1
05/01/1998	114916	1	712	5609	6	715	6036	8
05/01/1998	114938	1	713	200	1	715	6036	8
05/01/1998	115042	1	713	1200	2	715	6036	8
05/01/1998	115045	1	713	1000	1	715	6036	8
05/01/1998	115115	1	714	400	1	715	6036	8
05/01/1998	115143	1	713	1000	1	714	100	1
05/01/1998	115218	1	713	1000	1	715	6036	8
05/01/1998	115231	1	713	2019	2	715	6036	8
05/01/1998	115247	1	713	2019	2	715	6043	9
05/01/1998	115335	1	713	2019	2	714	50	1
05/01/1998	115349	1	713	2019	2	714	34	1
05/01/1998	115402	1	713	5109	3	714	34	1
05/01/1998	115426	1	713	2019	2	714	34	1
05/01/1998	115443	1	713	2019	2	714	15	1
05/01/1998	115448	1	713	1000	1	714	15	1
05/01/1998	115507	1	713	950	1	714	15	1

TABLE 3

Signed trades, Elf Aquitaine

Signed trades						
Day	Month	Second	Price	+/-	Duration	Buy/Sell
5	1	42276.0	714	1	0.5	0
5	1	42276.5	715	0	112.5	0
5	1	42389.0	713	0	93.0	1
5	1	42482.0	712	1	8.0	1
5	1	42490.0	713	1	0.5	0
5	1	42490.5	714	0	57.5	0
5	1	42548.0	712	1	8.0	1
5	1	42556.0	714	1	0.5	0
5	1	42556.5	715	0	88.5	0
5	1	42645.0	713	1	58.0	1
5	1	42703.0	714	0	204.0	1
5	1	42907.0	713	1	15.0	1

TABLE 4

Sell Trades, Elf Aquitaine

Sell trades					
Day	Month	Second	Price	+/-	Duration
5	1	42482.0	712	1	163.0
5	1	42645.0	713	1	58.0
5	1	42703.0	714	0	204.0
5	1	42907.0	713	1	23.0

Tables Legend

Table 1:

Date:	transaction date (format dd/mm/yyyy).
Time:	time (format : hhmmss).
Nb Seq.:	sequence number, <i>i.e.</i> execution order when there is simultaneous trading.
Price:	transaction price.
Quantity:	traded volume, <i>i.e.</i> number of traded shares.

Table 2:

Date:	transaction date format (dd/mm/yyyy).
Time:	transaction time (format: hhmmss).
Nb Seq.:	sequence number, <i>i.e.</i> execution order when there is simultaneous trading.
Price:	transaction price.
Bid/Ask:	best offered price to buy/sell.
Q^b/Q^a :	quantity proposed at the bid/ask.
Nb sellers/buyers:	number of buyers proposing the overall quantity Q^b /number of buyers proposing the overall quantity Q^a

Tables 3 and 4:

Day:	trading day
Month:	trading month.
Second:	transaction time in seconds, starting at 10 am (36000 seconds).
Price:	transaction price.
+/-:	indicator which equals 1, if the price increases right after the current trade, 0 otherwise.
Duration:	necessary time to have a price modification, starting at the current price.
Corrected Duration:	same as duration, but corrected from intraday seasonalities.
Buy/Sell:	indicator which equals to 1, if the trade is a sell, 0 otherwise.