CORE DISCUSSION PAPER
9944

TIME TRANSFORMATIONS, INTRADAY DATA AND VOLATILITY MODELS

Pierre Giot[1]

August 1999

Abstract

In this paper, we focus on the trade and quote data for the IBM stock traded at the NYSE. We present two different frameworks for analyzing this dataset. First, using regularly sampled observations, we characterize the intraday volatility of the mid-point of the bid-ask quotes by estimating GARCH and EGARCH models, with intraday seasonality being accounted for. We also highlight the impact of characteristics of the trade process (traded volume, number of trades and average volume per trade) on the volatility specifications. Secondly, we deal directly with the irregularly spaced data. We review two time transformations that allow a thinning of the original dataset such that new durations are defined. The newly defined price and volume durations are characterized and the performance of the Log-ACD model for modelling these durations is assessed. Moreover, price durations allow an easy computation of intraday volatility and this method compares favorably to ARCH estimations.

Keywords: intraday data, trades and quotes, intraday volatility, market liquidity

[1]CORE, Université Catholique de Louvain; email: giot@core.ucl.ac.be.

While remaining responsible for any errors in this paper, the author would like to thank Luc Bauwens, Joachim Grammig, Michel Lubrano and David Veredas for useful remarks and suggestions.
This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsability is assumed by the author.

1 Introduction

The recent availability of intraday financial databases[1] has had an important impact on research in applied econometrics and financial market microstructure. In the applied econometrics literature, this has given birth to the so-called high-frequency models, which attempt to describe characteristics of the price process (for example the volatility or the trading intensity) on an intraday basis. Broadly speaking, two main classes of high-frequency models exist. First, extensions of the ARCH type models that deal with regularly spaced data, and which focus on the volatility process during the day (Andersen and Bollerslev, 1997; Bollerslev and Domowitz, 1993). Secondly, duration models of the Autogressive Conditional Duration (ACD) type that accomodate irregularly spaced data[2].

These econometric developments have had important consequences, which go beyond the mere statistical description of the price process. With the available data and the newly developped models, it is now possible to apply (and test the relevance of) the models developped in the market microstructure literature. In this field, recent research has focused on the descriptive characteristics of the trade and bid-ask quote processes, the behaviour of market makers and the testing of models providing a theoretical framework of how market makers and traders interact in the trading process. Some key questions that are often investigated are: how do market makers fix their bid-ask quotes and the spread? Does the size of the spread depend on the information content of the trades? How quickly are quotes revised? How can the information content of a trade be measured (permanent effect of the trade/transitory effect)? See for example Easley and O'Hara (1992), Easley, Kiefer and O'Hara (1997) or Hasbrouck (1991) for some recent work in this field.

In this paper, our goal is to try to answer a relatively simple question: given an intraday tick-by-tick database for trades and bid-ask quotes[3], what are the possible ways of dealing with this new type of data? Although this sounds like a relatively simple task, it soon turns out that this type of analysis involves a lot of choices that have an impact on the analysis:

(i) is the data to be analyzed in a fixed interval framework or will models be built for the original irregularly spaced data? Working in a fixed interval frame-

[1] These intraday databases, also called tick-by-tick databases, are now available for most exchanges, such as the New York Stock Exchange, Paris Bourse, Chicago Mercantile Exchange, Brussels Stock Exchange, ... For currency trading, Olsen & Associates collected several years of data, which are available as the HFDF93 and HFDF96 databases.

[2] The original ACD model has been introduced by Engle and Russell (1998) and this model is the focus of ongoing research. See for example Bauwens and Giot (1997,1998), Bauwens and Veredas (1999), Grammig, Hujer, Kokot and Maurer (1998), Ghysels and Jasiak (1997) or Grammig and Wellner (1999).

[3] We address this issue for a stock traded on the NYSE. While much of the work presented here could be applied to financial assets traded on other exchanges, some features are particularly dependent on the information given in the dataset, e.g. the volume of the trades or the joint availability of trades and quotes.

work allows the use of "standard" time series tools such as the ARCH class models, but does not take into account the important information conveyed by the time between market events. Working with irregularly spaced data forbids using ARCH type models and calls for ACD type models. Moreover, if the data is to be dealt with in a fixed interval framework, time transformations are needed to convert the original irregularly spaced data into regularly spaced data. This usually involves sampling the data at a given frequency. What is the impact of the sampling frequency on the characteristics of the sampled process and estimated models? Furthermore, when working with regularly and irregularly spaced data, a strong intraday seasonality is exhibited by the data. This seasonality is dependent on characteristics of the exchange and the behaviour of market participants (opening and closing of trading, lunchtime, ...). This feature must be taken care of before using ARCH or ACD type models.

(ii) as the database gives information on both the trade and quote processes, how can these two processes be combined in the existing models? In a fixed interval framework, is information given by the trade process (such as the traded volume or number of trades for example) useful for characterizing properties of the quote process?

(iii) when dealing with irregularly spaced data, time (data) transformations can be applied to the data. Taking the trade and quote data as an input in a thinning algorithm[4], a new dataset can be defined which gives information on the volatility, the liquidity, the trading intensity, ... What are the characteristics of these newly defined marked point processes? Because these processes are irregularly spaced, they can also be modelled using the ACD class of models.

In this paper, we focus on these three points, with an emphasis on the interaction between the trade and quote processes, both for regularly and irregularly spaced data. GARCH and EGARCH models are estimated on the intraday *regularly sampled* quote data after taking into account the intraday seasonality. Traded volume, number of trades and average volume per trade are included in the EGARCH specification to assess the impact of trade related information on the conditional volatility of the quote process.

Next, we characterize several time transformations leading to newly defined *irregularly spaced* data that combine features of the trade and quote processes. Log-ACD models are estimated on these new datasets and the performance of the models are assessed. One of the main feature of this paper is to present in a single text the analysis of the same dataset in both the regularly spaced and

[4]By definition, thinning a marked point process (such as given by the trade or quote database) consists in retrieving a subset of the original dataset, such that particular features of the data are stressed (see Section 4).

irregularly spaced frameworks. Furthermore, intraday volatility is estimated using both an EGARCH model (on equidistantly sampled data) and a Log-ACD model (on irregularly spaced price durations), indicating that high frequency duration models provide an interesting alternative to ARCH type models.

Some of the issues mentionned above have already been addressed in the literature. For example, Le Fol and Mercier (1998) study several time deformation techniques in a fixed interval framework. Maillet and Michel (1998), Hafner (1996), Guillaume, Dacorogna, Davé, Muller, Olsen and Pictet (1995) or Guillaume, Dacorogna and Pictet (1997) focus on high frequency volatility models of the GARCH type applied to regularly spaced data for currency trading. Andersen and Bollerlev (1997) provide a thorough description of the application of GARCH models to high-frequency data (for currency trading at the FOREX and futures trading at the CME) and particularly insist on the need to take into account the intraday seasonal component prior to estimating the model. Regarding irregularly spaced data, Engle and Russell (1997, 1998) proposed the ACD model to explicitely model the time between market events. Bauwens and Giot (1997) introduce the Log-ACD model as an alternative to the ACD model and model the bid-ask quote process with respect to characteristics of the trade process.

The rest of the paper is organized in the following way. In Section 2, we introduce the dataset for the trades and quotes used in the empirical part of the paper. In Section 3, we deal with the regularly spaced data, estimate corresponding high frequency GARCH and EGARCH models for several sampling frequency and augment the quote dataset with features from the trade process. In Section 4, we work on the irregularly spaced data. We briefly review the ACD model of Engle and Russell (1998) and the Log-ACD model of Bauwens and Giot (1997). We introduce several time transformations that can be applied to the irregularly spaced data, characterize the properties of the resulting marked point processes and estimate Log-ACD models on the new datasets. We also estimate (as an alternative to the EGARCH model) the pattern of intraday volatility using a Log-ACD model and price durations. Section 5 concludes.

2 Information on the dataset

Recent econometric work on intraday data has usually focused on four databases, corresponding to four different exchanges: the NYSE and the Trade and Quote database (Bauwens and Giot, 1997,1998; Engle and Lunde, 1998) or the Trades, Orders Reports and Quotes database (Engle and Russell, 1998), the Olsen database for the FOREX (Bollerslev and Domowitz, 1993; Guillaume, Dacorogna, Davé, Muller, Olsen, and Pictet, 1995; Andersen and Bollerlev, 1997), the database of the Paris Bourse (Gouriéroux, Jasiak and Le Fol, 1996; Bisière and Kamionka, 1998; Darolles, Gouriéroux and Le Fol, 1998) and the database of the German Stock Exchange (Grammig, Huher, Kokot, and Maurer, 1998; Grammig and Wellner,

1999). In this paper, we use the Trade and Quote database[5], which is the official database released by the NYSE.

The NYSE is the largest stock exchange in the United States. It is more than 200 years old and the total market capitalization of its listed stocks is close to 12 trillion dollars, with the average volume of shares traded in one day greater than 600 million shares. Trading at the NYSE is based on a so-called hybrid system, i.e., it uses a trading mechanism combining a market maker system and an order book system. Assigned to each stock is one market maker (called the specialist), who is making the market for this stock. His main tasks are:

- manage the trading and quoting processes (i.e., ensure an orderly market, report quotes and trades, ...);

- provide liquidity when necessary, by taking the other side of the trades when needed. Most of the trades are however executed against standing orders in the order book, or they go through the floor traders.

Thus, the specialist monitors *all* the trading process, which takes place from 9h30 to 16h. Apart from an opening auction, trading is continuous.

Retrieving the data for the IBM stock over the period September 96 to November 96, we construct two separate databases, one for the trades and the other for the bid-ask quotes. Our main focus is on the bid-ask quote database as we wish to characterize the bid-ask quote process. However, as detailed in Section 4, the trade database is needed as thinning the data for the bid-ask quotes requires taking into account information given by the trade process. Because all records listed in the TAQ database are not valid[6], we select the "regular" trade and bid-ask quotes recorded between 9:30 AM and 16:00 PM.

With respect to this dataset, let us define a marked point process for the bid-ask quotes Q and a second one for the trades P. Q is defined as a marked point process referenced by t_i, where t_i indicates the time in seconds when the bid-ask quotes are posted by the specialist. When posting the quotes, the specialist specifies[7] the current bid and ask prices b_i and a_i. Thus, the marked point process for the quotes is (t_i, b_i, a_i), for $i = 1 \dots n$, where n is the total number of quotes.

Regarding the trade process, the marked point process is defined by (y_j, p_j, v_j), for $j = 1 \dots m$, where m is the total number of trades and y_j is the time in seconds when the trade was made, p_j designates the price of the trade and v_j the corresponding volume. Merging both databases and using a signing algorithm as proposed by Lee and Ready (1992), the marked point process for the trades is

[5]This database consists of two parts: the first reports all trades, while the second lists the bid and ask prices posted by the specialists (on the NYSE).

[6]All trades have a correction and a condition indicator, giving information on the validity of the trade.

[7]These bid and ask prices are valid for a certain number of shares, called the depth at the bid or at the ask. This information is also given by the specialist, but we do not include it in our dataset.

extended to $(y_j, p_j, v_j, s_j, b_j, a_j)$ where s_j indicates if the trade is a buy or a sell, while b_j and a_j indicate the existing bid and ask quotes when the trade was made.

By definition, the trade and bid-ask quote datasets contain irregularly time spaced data as trades and quotes are recorded as soon as they are reported. Thus the durations between two trades $z_j = y_j - y_{j-1}$ and between two bid-ask quotes $x_i = t_i - t_{i-1}$ are not constant. Table 1 gives some descriptive information on both series. In the next sections, we investigate two different ways of dealing with this dataset.

Table 1: Information on the trade and quote processes (IBM stock)

	Trades	Quotes
Number	60,718	33,185
Mean duration	24	43.9
Standard dev. of durations	34.6	65
Minimum duration	0	1
Maximum duration	674	1221

Data extracted from the September, October and November 1996 TAQ CD-ROM. All numbers (except the number of trades and quotes) are in seconds.

3 Regularly spaced data

As introduced in the last section, the trade and bid-ask quote marked point processes are irregularly spaced in time. In this section, we present a first way of looking at this dataset, i.e., by working in a regularly time spaced framework using a resampled bid-ask quote dataset based on the previously defined Q. Our aim is to characterize the properties of the returns defined on the new regularly spaced bid-ask quotes and estimate volatility models of the GARCH and EGARCH class on the dataset. We also highlight the impact of the traded volume, number of trades and average volume per trade on the volatility specifications. The methodology of the work presented in this section is close to Andersen and Bollerslev (1997), which focused on high-frequency data for the FOREX and CME[8].

[8]However, we do not consider the aggregation properties of the estimated GARCH models, but we stress the interaction between the trade and bid-ask quote processes.

3.1 Transaction time

When dealing with intraday data, a first possibility is to treat the marked point process as a collection of numbered observations, thus "forgetting" the information given by the time between the obervations and overlooking the fact that they are not regularly spaced in time. This defines the "intraday" returns[9] as $r_i = ln(p_i) - ln(p_{i-1})$, which are also said to be defined in transaction time. If these returns are directly used in a standard time series model, it can lead to meaningless estimation results as it treats all observations as being equidistantly spaced in time, which they are not. Furthermore, transaction time assumes that all observations convey the same information, whatever their spacing in time and associated characteristics such as volume. With these drawbacks, it is not suprising that transaction time returns are not much used in the empirical literature.

3.2 Equidistant sampling

Time transformations that can be applied to intraday data have been recently studied in the literature. For example, Le Fol and Mercier (1998) review most of the possible time transformations that can be used. They highlight the characteristics of several time scales (real time, transaction time, volume time, ...), including the time transformations introduced by Olsen & Associates to take into account the intraday seasonality in high frequency FOREX data. Maillet and Michel (1998) focus on stock returns distribution in a volume time scale, i.e., a time where transaction volume is constant between observations. For the same dataset, they compare the empirical characteristics of returns both regularly sampled and sampled in a volume time scale. In this subsection, we focus on equidistantly sampled data[10], with an emphasis on the estimation of intraday volatility using GARCH and EGARCH models.

To transform the original dataset in a new one featuring regularly spaced observations with respect to time, we first define a sampling grid with an associated sampling time equal to h seconds. As we focus on data released by the NYSE, the sampling grid is defined from 9h30 to 16h. The new marked point process Q' is defined as $(t_i^{'}, b_i^{'}, a_i^{'})$, where the $t_i^{'}$ define the grid over the period 9h30-16h with a h long sampling interval, i.e., that $t_i^{'} - t_{i-1}^{'}$ is equal to h for all i. Associated to the $t_i^{'}$ are the $b_i^{'}$ and $a_i^{'}$, the sampled bid and ask quotes. For all $t_i^{'}$, they are the most recently recorded b_i and a_i, except for the first observation of the day which consists of the first b_i and a_i. The return process is then defined on the mid-point of the bid and ask quotes, i.e., the return process consists of all $r_i = ln(p_i^{'}) - ln(p_{i-1}^{'})$, where $p_i^{'} = (a_i^{'} + b_i^{'})/2$. Because we wish to focus on the intraday returns, we delete the first return of the day as it is the overnight return, i.e., the return between the

[9] When working with bid and ask quotes, the returns are usually defined on the mid-point, defined as $p_i = (b_i + a_i)/2$.

[10] We choose to focus on "straightforward" equidistant sampling, i.e., we do not consider the different time scales introduced by Le Fol and Mercier (1998) before sampling our data.

last recorded price before 16h30 and the first recorded price after 9h30.

Table 2: Equidistant sampling

Information on the returns

	h=1 min	h=5 min	h=10 min	h=15 min	h=30 min
Oversampling	33.5	4.1	0.8	0.06	0
Q(10)	146.9	51.5	50.6	32.3	19.7
Kurtosis	34.1	13.9	15.3	14.6	14.8
Kurtosis'	10.9	7.3	6.2	5.7	7.5
Observations	24389	4877	2438	1625	812

Information on the squared returns

	h=1 min	h=5 min	h=10 min	h=15 min	h=30 min
Q(10)	3162.9	778.8	401.3	364.1	35.7
Q(10)'	3062.3	732.4	492.2	340.4	32.1
$\rho_1^{'}$	0.19	0.18	0.20	0.35	0.10
$\rho_2^{'}$	0.13	0.14	0.16	0.22	0.08

Descriptive statistics for the equidistantly sampled (intraday) returns and squared returns for the IBM stock traded at the NYSE. The period under review is September 96 - November 96 or 13 weeks. $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the returns and squared returns. Q(10)' denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the square of the intraday seasonally adjusted returns. Kurtosis denotes the coefficient of kurtosis on the returns, while Kurtosis' denotes the coefficient of kurtosis on the intraday seasonally adjusted returns. $\rho_1^{'}$ and $\rho_2^{'}$ denote the first two autocorrelation on the square of the intraday seasonally adjusted returns. The intraday seasonally adjusted returns are defined by dividing the raw returns by the square root of the deterministic volatility function. The oversampling rate is defined as the ratio between the identically sampled quotes and all sampled quotes.

As indicated in Table 2 which gives descriptive statistics for the returns and squared returns, choosing a high sampling frequency (for example $h = 1$ minute) leads to a very large oversampling rate[11], i.e., sampled quotes are often identical to the previously sampled ones as no new information has been released during the sampling interval h. As a result, the Q statistic for the returns is very large, indicating that these returns are (spuriously) correlated. Thus, depending on the

[11]The oversampling rate is defined as the ratio between the identically sampled quotes and all sampled quotes.

trading activity of the stock, h should be chosen small enough to get a detailed picture of intraday volatility (see below), but not too small so that oversampling is limited. Table 2 also indicates that the intraday returns feature a large kurtosis (around 14), and that the squared returns are strongly autocorrelated, with a very large value for the Q-statistic, indicating a strong persistence for the intraday volatility.

3.3 Intraday volatility and seasonality

Regarding the reporting of trades and bid-ask quotes, it is well known that the trading activity is not constant along the trading day. More trades and bid-ask quotes are recorded in the early and late trading hours rather than around lunch time. The intraday seasonality observed for the trading and quoting activity has also been found to hold for other market characteristics such as the volatility or the spread. See for example Brock and Kleidon (1992) or Engle and Russell (1998) for the NYSE and Bollerslev and Domowitz (1993) or Andersen and Bollerslev (1997) for the FOREX.

In order to estimate the pattern of intraday volatility for the equidistantly space returns defined above, several possibilities arise. In the empirical market microstructure literature which usually focuses on the deterministic pattern of intraday volatility, it is often computed as the variance of returns in a given time interval, or as the average of the squared returns during a time interval (see for example Ederington and Lee, 1993, or Gwilym, Buckle, and Thomas, 1997). For example, the average pattern of intraday volatility could be defined by computing averages (over 30 minute intervals for example) of squared returns defined on the mid-point of the bid-ask quotes recorded during these intervals.

A more interesting way of estimating intraday volatility when dealing with regularly spaced data is to combine the computation of average intraday volatility patterns with the use of volatility models of the ARCH type for example[12]. This procedure involves

(i) the computation of intraday volatility patterns, defined as the intraday seasonal component of the volatility;

(ii) the computation of the deseasonalized volatility, i.e., the volatility from which the seasonal component has been removed, usually in a multiplicative way;

(iii) the estimation of ARCH type models on the deseasonalized volatility if necessary, i.e., if the seasonal component does not remove all existing patterns.

[12] A recent example of this procedure which is applied to FOREX and futures data is given in Andersen and Bollerslev (1997). In this paper, the two authors particularly stress the danger of estimating ARCH type models on the high-frequency data without removing the deterministic intraday pattern beforehand.

The importance of taking into account the intraday seasonal component of the volatility can be seen by plotting an autocorrelogram for the squared returns. Working with 5 and 10 minute sampling intervals, Figure 1a and 1c give the autocorrelograms for the first 200 lags computed on the raw intraday returns. With a 5 (10) minute long sampling interval, there are 78 (39) intervals during one trading day. In Figure 1a (1c), the volatility peaks at lags multiple of 78 (39) are clearly visible.

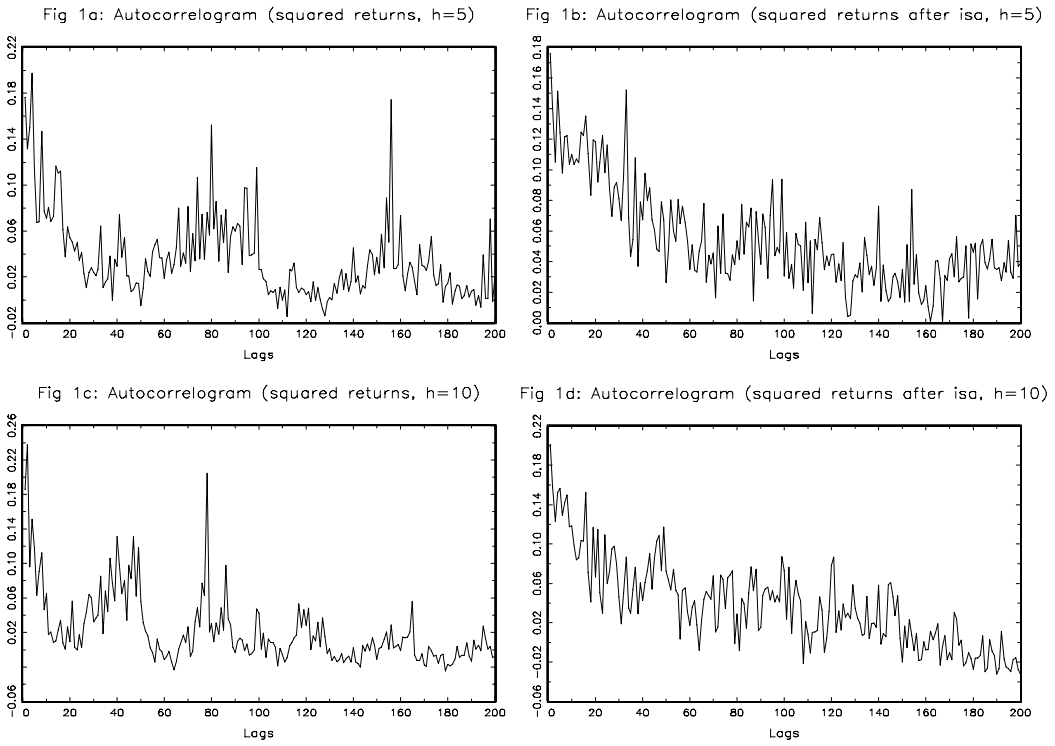

Figure 1: Autocorrelogram for the equidistantly sampled returns (h=5 and 10 minutes), before (Figures 1a and 1c) and after (Figures 1b and 1d) intraday seasonal adjustment (isa). The returns are defined on the mid-point of the bid-ask quotes of the IBM stock. The period under study is September 96 - November 96 or 13 weeks.

To compute the intraday seasonal component of the volatility, we first define 30 minute intervals in which we compute the average squared returns defined on the mid-point of the sampled bid-ask quotes recorded during these intervals. This rather crude pattern of intraday volatility is then smoothed by using cubic splines and the resulting deterministic intraday volatility function is called $\phi(t_i)$. Figure 2 gives the intraday volatility patterns for the returns sampled with $h = 10$ minutes.

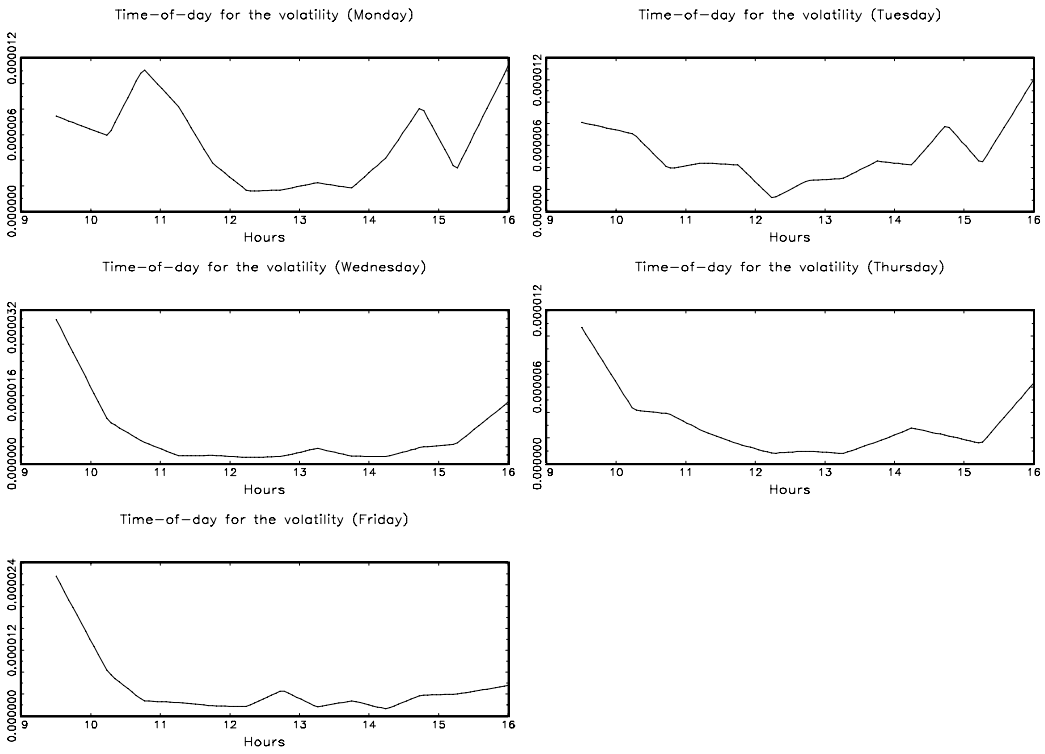

Figure 2: Intraday seasonal component for the volatility computed on equidistantly sampled returns (h=10 minutes). The returns are defined on the mid-point of the bid-ask quotes of the IBM stock. The period under study is September 96 - November 96 or 13 weeks.

In a second step, the deseasonalized returns are computed by dividing the raw returns by the square root of the deterministic pattern of intraday volatility, i.e., dividing the r_i by the square root of $\phi(t_i)$. Figure 1b (for h=5 minutes) and 1d (for h=10 minutes) give the first 200 lags of the autocorrelogram computed for the deseasonalized returns. As indicated by these autocorrelograms, the intraday seasonal peaks are no longer visible. However, from Figures 1b, 1d and the Q statistics for the squared deseasonalized returns given in Table 2, it is obvious that taking into account the intraday seasonality has not removed the autocorrelation in the volatility. This suggests that the remaining autocorrelation should be taken care of with a volatility model of the ARCH type for example.

In this section, we estimate two specifications of the ARCH type on the deseasonalized returns sampled at different intraday frequencies. First we model the conditional volatility as being a GARCH(1,1) process[13], and then we introduce an asymmetric effect by using an EGARCH(1,1) model[14]. As the results in Table 2 indicate that the returns are slightly correlated, an AR(1) structure is fitted on the returns. Thus the two specifications are:

(i) GARCH(1,1)

$$r_t = \beta_0 + \beta_1 r_{t-1} + \epsilon_t \tag{1}$$

with the error term modelled as

$$\epsilon_t = \sqrt{g_t}\eta_t \tag{2}$$

where η_t is assumed to be an independent Gaussian process, and the conditional volatility g_t is defined as

$$g_t = \kappa + \alpha\epsilon_{t-1}^2 + \delta g_{t-1} \tag{3}$$

with κ, α and $\delta > 0$ and $\alpha + \delta < 1$.

(ii) EGARCH(1,1)

$$r_t = \beta_0 + \beta_1 r_{t-1} + \epsilon_t \tag{4}$$

with the error term modelled as

$$\epsilon_t = \sqrt{g_t}\eta_t \tag{5}$$

where η_t is assumed to be an independent Gaussian process, and the conditional volatility g_t is defined as

[13]The GARCH model was introduced by Bollerslev (1986).

[14]The EGARCH model was introduced by Nelson (1991), as a way of modelling the asymmetric response of volatility.

$$ln(g_t) = \kappa + \theta\eta_{t-1} + \gamma(|\eta_{t-1}| - \sqrt{2/\pi}) + \delta ln(g_{t-1}) \quad (6)$$

where $\delta < 1$.

Table 3: Equidistant sampling

GARCH(1,1)

	h=5 min	h=10 min	h=15 min	h=30 min
β_0	0.033 (0.012)	0.046 (0.017)	0.061 (0.022)	0.091 (0.035)
β_1	-0.040 (0.016)	0.016 (0.023)	0.063 (0.027)	0.088 (0.038)
κ	0.020 (0.004)	0.043 (0.012)	0.037 (0.012)	0.041 (0.014)
α	0.068 (0.008)	0.108 (0.019)	0.079 (0.016)	0.057 (0.016)
δ	0.912 (0.010)	0.848 (0.028)	0.882 (0.026)	0.910 (0.023)
$Q(10)$	2.29	10.79	9.46	1.14

EGARCH(1,1)

	h=5 min	h=10 min	h=15 min	h=30 min
β_0	0.046 (0.012)	0.054 (0.017)	0.074 (0.022)	0.107 (0.035)
β_1	-0.043 (0.015)	0.019 (0.022)	0.063 (0.027)	0.069 (0.038)
κ	0.007 (0.002)	0.004 (0.004)	0.003 (0.004)	0.015 (0.006)
θ	0.029 (0.006)	0.033 (0.012)	0.038 (0.014)	0.005 (0.023)
γ	0.137 (0.013)	0.213 (0.027)	0.161 (0.030)	0.168 (0.033)
δ	0.975 (0.004)	0.952 (0.010)	0.955 (0.013)	0.964 (0.013)
$Q(10)$	6.75	11.81	11.12	1.98

Estimation results for the GARCH model as given by (1)-(3) and the EGARCH model (4)-(6). Both models are applied to 13 weeks of intraday data for the IBM stock. $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the square of the standardized returns η_t in (5).

The EGARCH(1,1) model allows for an asymmetric response of g_t to volatility shocks on the error terms ϵ_t or η_t. When η_{t-1} is negative, the slope of the conditional log-variance is equal to $\theta - \gamma$ and is equal to $\theta + \gamma$ when the standardized innovation is positive. Empirical studies conducted on daily data using EGARCH specifications for the conditional log-variance usually conclude that negative shocks have a more ronounced impact on volatility, i.e., that γ is negative (Nelson, 1991).

The results given in Table 3 indicate that

- both the GARCH(1,1) and the EGARCH(1,1) models are quite successful in taking into account the autocorrelation in the volatility of the deseasonalized returns. For all models and for all sampling intervals, the Q(10) statistics on the standardized residuals η_t are not significant at the five percent level.

- the GARCH effect is quite large, with $\alpha + \delta$ close to 0.9 in the GARCH specification, and the δ close to 0.95 in the EGARCH model.

- the asymmetric effect is positive and small in the EGARCH specification. For example, with $h = 10$ minutes and η_{t-1} negative, the slope of the conditional variance is equal to -0.18; it is equal to 0.246 when the standardized innovation is positive.

- at the shortest time interval considered here ($h = 5$ minutes), β_1 is significantly negative in both specifications, while it is positive for the other sampling frequencies. This slightly negative autocorrelation at the highest frequency indicates a reverting effect for the corresponding returns.

Combining the conditional volatility given by the GARCH or EGARCH models and the deterministic intraday patterns computed using the cubic splines, it is possible to obtain a precise pattern of intraday volatility for the raw returns. If g_t denotes the conditional variance of the deseasonnalized returns and $\phi(t)$ is the deterministic pattern of intraday volatility at time t, the conditional variance of the returns is then given by $g_t\phi(t)$. For the returns sampled at $h = 10$ minutes and using the fitted volatility function from the EGARCH specification, Figure 3 plots the annualized square root[15] of $g_t\phi(t)$ and the annualized absolute returns for the second week of September 96. As indicated in Figure 3, the conditional volatility tracks quite well the observed volatility.

3.4 Augmenting the dataset with trade related information

In the literature on GARCH models, volume has been found to have a high impact on the estimated coefficients of the model when included in the specification of the conditional variance. Using the theoretical framework provided by the so-called mixture of distributions hypothesis, Lamoureux and Lastrapes (1992) empirically verify that including volume as an additional variable in the conditional variance equation leads to a very significant decrease in the autoregressive coefficient of the variance equation, i.e., $\alpha + \beta$ of the GARCH process is close to zero. In their

[15]The annualized square root of the conditional variance has a more intuitive meaning than $g_t\phi(t)$. It is computed as $\sqrt{g_t\phi(t)250(16-9.5)3600/h}$, as it is assumed that there are 250 trading days and (16-9.5) $3600/h$ sampling intervals (h is expressed in seconds).

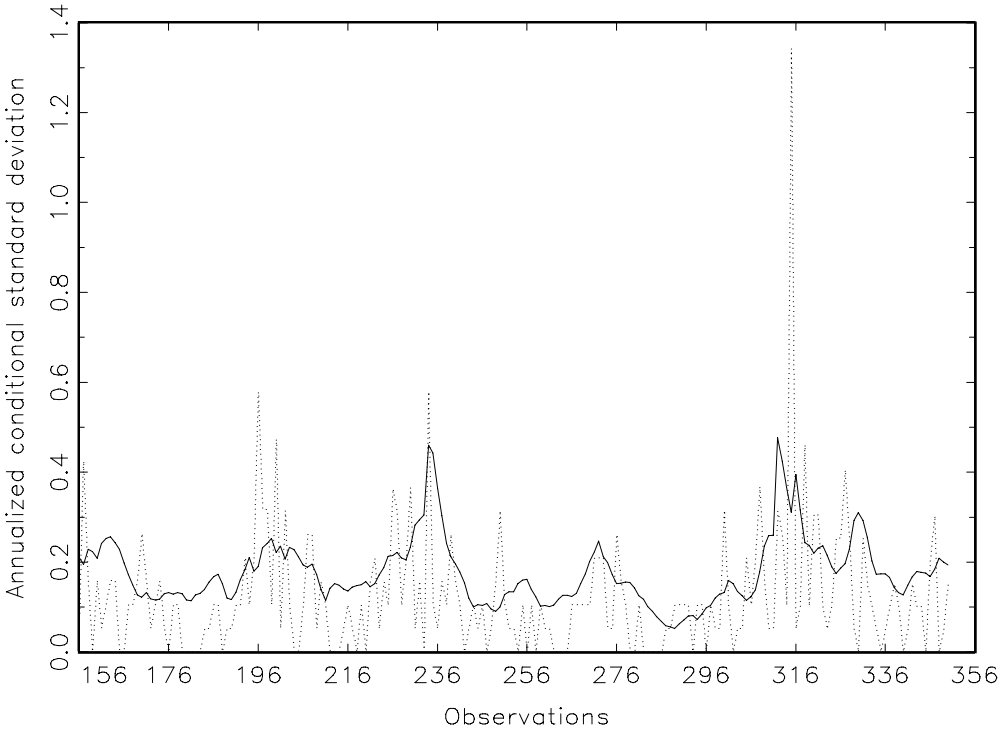

Figure 3: Annualized conditional standard deviation as forecasted by the EGARCH model (continuous line) and observed absolute returns. The returns are defined on the mid-point of the bid-ask quotes of the IBM stock (h=10 minutes). The model is estimated for the September 96 - October 96 period and the results are plotted for the second week of September.

detailed study based on daily data for US stocks, the GARCH effect completely disappears once volume is included in the conditional variance equation.

Using the available tick-by-tick data, this specification can also be tested, which enables us to assess the link between volume and volatility on an intraday basis. Merging the trade and bid-ask quote datasets, the marked point process for the equidistantly sampled bid-ask quotes Q' becomes $(t_i^{'}, b_i^{'}, a_i^{'}, v_i^{'})$, where $v_i^{'}$ is the total traded volume between $t_{i-1}^{'}$ and $t_i^{'}$, i.e., the volume traded during h seconds. The first and second columns in Table 4 report the estimated coeffcients of the EGARCH(1,1) model[16] defined previously with the traded volume added as an additional explicative variable in the conditional variance equation. The specification for the conditional log-variance is

$$ln(h_t) = \kappa + \theta\eta_{t-1} + \gamma(|\eta_{t-1}| - \sqrt{2/\pi}) + \delta ln(h_{t-1}) + \zeta_1 v_t \qquad (7)$$

where v_t is the total traded volume in the h second long period preceding t. Because the traded volume exhibits a strong intraday pattern (see Figure 4), it is first deseasonalized using the already defined cubic splines function.

As indicated in Table 4, the coefficient δ decreases from a value close to 0.95 to almost zero once the traded volume is included in the EGARCH specification for the conditional variance. These results are in agreement with the previously documented results by Lamoureux and Lastrapes (1992) on daily data and Maillet and Michel (1998) for intraday data on Elf-Aquitaine (a large liquid stock traded at the Paris Bourse). Thus, traded volume seems to affect the conditional volatility of the price process in the same way, whether one works with daily or intradaily data.

Regarding the volatility-volume relationship, Jones, Kaul and Lipson (1994) suggest that it is the number of trades, and not the average volume per trade, that is the main driving force behind volatility. In the Lamoureux and Lastrapes framework, both effects are intertwined as the traded volume is the product of the number of trades and the average volume per trade. Jones, Kaul and Lipson (1994) "split" the traded volume into two components, the number of trades and the average volume per trade, which they use as additional variables in a volatility equation[17]. Indeed, using daily data for a large number of US stocks traded on the NASDAQ, they show that the volatility-volume relationship disappears when the number of trades is included in the regression.

On an *intraday* basis, the impact of the average volume per trade and the number of trades on the volatility equation can easily be assessed using the merged databases. First, Q' is augmented to $(t_i^{'}, b_i^{'}, a_i^{'}, v_i^{'}, av_i^{'}, n_i^{'})$, where $av_i^{'}$ is the average volume per trade and $n_i^{'}$ is the number of trades between $t_{i-1}^{'}$ and $t_i^{'}$. Combining

[16] Similar tests were conducted on the GARCH specification. The results are identical and are not reported here.

[17] While Lamoureux and Lastrapes (1992) use a GARCH framework to test their model, Jones, Kaul and Lipson (1994) "simply" regress the daily volatility on the average volume per trade and the number of trades.

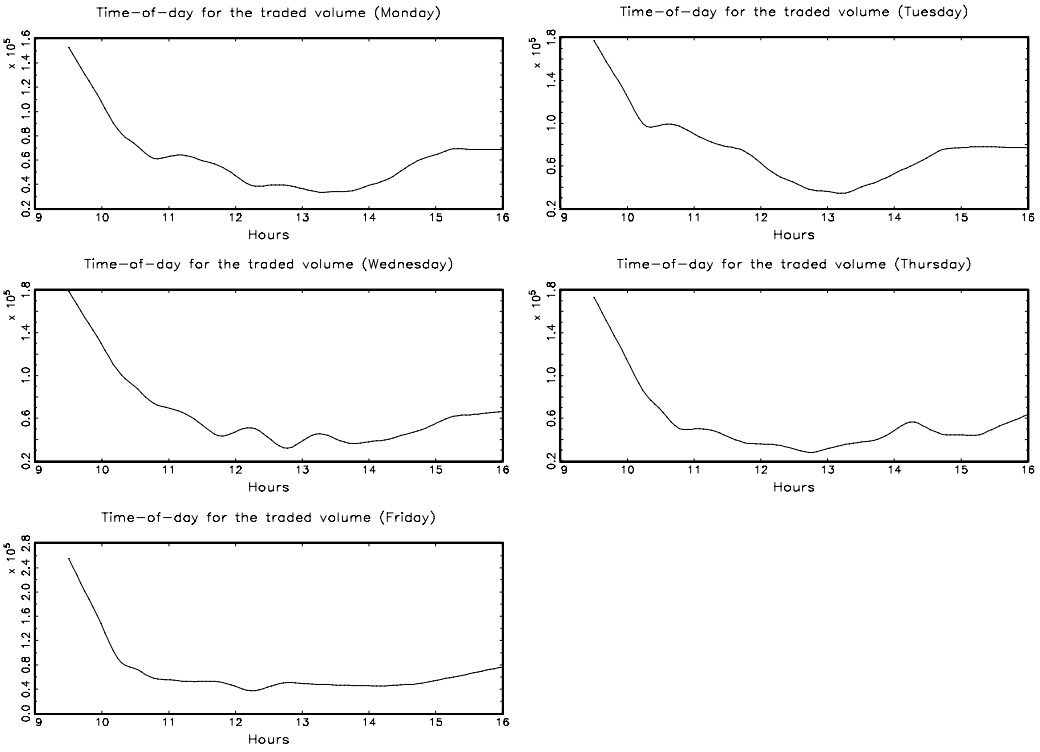

Figure 4: Intraday seasonal component for the traded volume (h=10 minutes), defined as the total traded volume during 10 minute long intervals.

Table 4: Equidistant sampling
EGARCH(1,1) with additional variables
(traded volume and number of trades)

	h=5 min	h=10 min	h=5 min	h=10 min
β_0	-0.014 (0.010)	-0.006 (0.014)	-0.007 (0.010)	0.001 (0.014)
β_1	-0.097 (0.014)	-0.066 (0.019)	-0.098 (0.013)	-0.068 (0.020)
κ	-1.233 (0.042)	-1.457 (0.055)	-1.907 (0.044)	-1.967 (0.069)
θ	-0.016 (0.021)	-0.076 (0.029)	0.001 (0.020)	-0.080 (0.029)
γ	0.134 (0.029)	0.135 (0.044)	0.041 (0.029)	0.059 (0.044)
δ	0.106 (0.027)	0 (.)	0 (.)	0 (.)
ζ_1	0.880 (0.031)	1.051 (0.046)	-	-
ζ_2	-	-	1.470 (0.039)	1.569 (0.063)
$Q(10)$	96.76	21.55	62.44	24.25

Estimation results for the EGARCH models as given by (4)-(7) and (4)-(8) applied to 13 weeks of intraday data for the IBM stock. ζ_1 is the coefficient of the traded volume in (7) and ζ_2 is the coefficient of the number of trades in (8). $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the square of the standardized returns η_t in (5).

the previously defined EGARCH model with the additional explicative variables[18] n_t (n_t is the number of trades in the h second long period preceding t) and av_t (av_t is the average volume per trade in the h second long period preceding t) defines the following EGARCH models:

$$ln(h_t) = \kappa + \theta\eta_{t-1} + \gamma(|\eta_{t-1}| - \sqrt{2/\pi}) + \delta ln(h_{t-1}) + \zeta_2 n_t \tag{8}$$

$$ln(h_t) = \kappa + \theta\eta_{t-1} + \gamma(|\eta_{t-1}| - \sqrt{2/\pi}) + \delta ln(h_{t-1}) + \zeta_3 av_t \tag{9}$$

$$ln(h_t) = \kappa + \theta\eta_{t-1} + \gamma(|\eta_{t-1}| - \sqrt{2/\pi}) + \delta ln(h_{t-1}) + \zeta_4 av_t + \zeta_5 n_t \tag{10}$$

The estimation results are reported in the right part of Table 4 and Table 5. As indicated by Table 4, including the number of trades or the volume traded in

[18]Like the traded volume, the number of trades and the average volume per trade display a strong intraday seasonality, which is removed from the data using the procedure defined above for the volatility.

Table 5: Equidistant sampling
EGARCH(1,1) with additional variables
(average volume per trade and number of trades)

	h=5 min	h=10 min	h=5 min	h=10 min
β_0	0.040 (0.012)	0.040 (0.016)	-0.018 (0.009)	-0.008 (0.014)
β_1	-0.040 (0.015)	0.015 (0.023)	-0.105 (0.013)	-0.078 (0.019)
κ	-0.096 (0.016)	-0.381 (0.074)	-2.328 (0.060)	-2.606 (0.089)
θ	0.020 (0.008)	-0.010 (0.022)	-0.007 (0.020)	-0.087 (0.029)
γ	0.161 (0.015)	0.328 (0.036)	0.073 (0.029)	0.084 (0.044)
δ	0.948 (0.008)	0.808 (0.039)	0.004 (0.025)	0 (.)
ζ_3	0.098 (0.015)	0.360 (0.068)	-	-
ζ_4	-	-	0.445 (0.034)	0.655 (0.061)
ζ_5	-	-	1.405 (0.045)	1.498 (0.066)
$Q(10)$	3.96	16.67	84.69	24.42

Estimation results for the EGARCH models as given by (4)-(9) and (4)-(10) applied to 13 weeks of intraday data for the IBM stock. ζ_3 is the coefficient of the average volume per trade in (9), ζ_4 is the coefficient of the average volume per trade in (10) and ζ_5 is the coefficient of the number of trades in (10). $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the square of the standardized returns η_t in (5).

the specification for the conditional variance has a similar effect on the estimated coefficients of the EGARCH process, as in both cases the autoregressive effect disappears completely. The results given in Table 5 indicate that the average volume per trade has a relatively small impact on the coefficients of the volatility equation. For $h = 5$ minutes, δ decreases from 0.975 to 0.948 and for $h = 10$ minutes, it decreases from 0.952 to 0.808. However, as soon as the number of trades is included as an extra variable, the coefficient δ goes down to zero and we get the same results as discussed above. Thus the results reported in Table 4-5 indicate that, out of the three volume related variables, it is the number of trades that matters most when conditional volatility is modelled.

4 Irregularly spaced data

In Section 3, the dataset was transformed using equidistant sampling so that "standard" time series techniques could be used on the data. While allowing the use of the now well documented time series techniques of the ARCH class, this time transformation removes the time between market events from the dataset and thus loses important information. In the financial literature dealing with the intraday characteristics of an asset (price process, liquidity, behaviour of market agents trading the asset), time has long been considered as being exogenous, with the implication that time between market events does not matter. See for example Kyle (1985) or Glosten and Milgrom (1985). For a broad range of empirical studies, such as Glosten and Harris (1988) or Hasbrouck (1991), the time between market events does not enter the analysis either.

However, in the recent market microstructure literature[19], the role of time has been found to be of particular importance when the behaviour of market agents has to be modelled. For example, in the Easley and O'Hara (1992) model, time is no longer exogenous but has a deep impact on the way market makers update their quotes. More precisely, an active market, i.e., a market featuring short durations between trades, is usually associated with possible informed trading and leads to more frequent bid-ask quote updates by the market maker (and also to an increase in the quoted spread).

From an econometric point of view, this literature has provided much support for econometric models dealing with the time between market events. When it comes to dealing with the time between market events, the most well-known model is the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998). Over the last four years, this model has attracted a growing interest as new models based on the ACD specification have been developped[20].

While the literature on possible models is growing fastly, there is surprisingly little concern about the different durations between specific market events that can be defined. For example, it is possible to define durations as the time between trades, the time between quotes, the time between quotes leading to a given price change, the time between quotes such that a given volume is traded, ... In all cases, the focus is set on durations but the implications for the economic framework are different in each case.

In this section, we first briefly review the basic high frequency duration models (ACD and Log-ACD) that are now commonly used. Secondly, we characterize the possible time transformations that can be applied to the previously defined dataset. For each transformation, we give the precise meaning of the durations that are being modelled and we look at the characteristics of the estimated models.

[19] O'Hara (1995), Biais, Foucault and Hillion (1997) and Goodhart and O'Hara (1997) are excellent surveys of the existing theoretical and empirical models developed in this field.

[20] See for example Bauwens and Giot (1997, 1998), Bauwens and Veredas (1999), Grammig, Hujer, Kokot and Maurer (1998), Ghysels and Jasiak (1997).

4.1 The ACD and Log-ACD models

Let x_i be the duration between two market events that happened at times t_{i-1} and t_i, i.e., $x_i = t_i - t_{i-1}$. The assumption introduced by Engle and Russell (1998) is that the time dependence in the durations can be subsumed in their conditional expectations $\Psi_i = E(x_i|I_{i-1})$, in such a way that x_i/Ψ_i is independent and identically distributed. I_{i-1} denotes the information set available at time t_{i-1}, supposed to contain at least $\tilde{x}_{i-1}$ and $\tilde{\psi}_{i-1}$, where $\tilde{x}_{i-1}$ denotes x_{i-1} and its past values, and likewise for $\tilde{\psi}_{i-1}$. The ACD model specifies the observed duration as a mixing process:

$$x_i = \Phi_i \epsilon_i \tag{11}$$

where the ϵ_i are IID and follow a Weibull(1,γ) probability distribution, while the Φ_i are proportional to the conditional expectation of x_i as explained below.

A second equation specifies an autoregressive model for the (expected) conditional durations:[21]

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1} \tag{12}$$

with the following constraints on the coefficients: $\omega > 0$, $\beta \geq 0$, $\alpha \geq 0$ and $\alpha + \beta < 1$. As $\Psi_i = E(x_i|I_{i-1})$, we have that

$$\Psi_i = \Gamma(1 + \frac{1}{\gamma})\Phi_i \tag{13}$$

where $\Gamma(.)$ is the gamma function[22]. Because the ϵ_i are Weibull(1,γ), the conditional hazard for x_i is given by

$$h(x_i|I_{i-1}) = \frac{\gamma}{\Psi_i} \left(\frac{x_i}{\Psi_i}\right)^{\gamma-1} \tag{14}$$

As shown in Engle and Russell (1998), the above defined ACD model can account for a clustering effect on the durations and overdispersion. Indeed, the autoregressive structure on the conditional expectation of the durations implies that small durations are more likely to be followed by small durations (and likewise for long durations).

With respect to the ACD model defined above, the logarithmic version of the ACD model[23] changes the mixing process (11) into the following equation:

$$x_i = e^{\phi_i} \epsilon_i \tag{15}$$

where the ϵ_i are IID and follow a Weibull(1,γ) distribution, while ϕ_i is proportional to the logarithm of the conditional expectation of x_i as explained below.

[21] This model is the ACD(1,1). More lags of x_i and Ψ_i can be added.

[22] If $\gamma = 1$, the Weibull distribution becomes an exponential one. In this case, $\Phi_i = \Psi_i$.

[23] The Log-ACD model was introduced by Bauwens and Giot (1997).

Let ψ_i be the logarithm of the conditional expectation of x_i, so that $\psi_i = \ln E(x_i|I_{i-1})$. A second equation specifies an autoregressive model for the logarithm of the conditional durations:

$$\psi_i = \omega + \alpha g(x_{i-1}, \epsilon_{i-1}) + \beta\psi_{i-1} \tag{16}$$

For positivity of e^{ψ_i} and thus of x_i, there are no restrictions on the sign of the parameters ω, α and β. As $\psi_i = \ln E(x_i|I_{i-1})$, we have that

$$e^{\phi_i}\Gamma(1+1/\gamma) = e^{\psi_i} \tag{17}$$

While Bauwens and Giot (1997) propose several choices of $g(x_{i-1}, \epsilon_{i-1})$, the most interesting specification seems to be when $g(x_{i-1}, \epsilon_{i-1}) = \epsilon_{i-1}$. With this choice of g, equation (16) can be written as

$$\psi_i = \omega + \alpha\epsilon_{i-1} + \beta\psi_{i-1} = \omega + \alpha\frac{x_{i-1}\Gamma(1+1/\gamma)}{e^{\psi_{i-1}}} + \beta\psi_{i-1} \tag{18}$$

With this specification, the logarithm of the conditional expectation depends on its past lagged value and on the lagged "excess duration". This model is close to the exponential GARCH model of Nelson (1991). For covariance stationarity of ψ_i, $|\beta|$ must be smaller than one.

4.2 Time transformations

The ACD and Log-ACD models are powerful tools when times between market events are to be modelled. They can easily take into account the main characteristics of the data (overdispersion, clustering of the durations) and they provide a convenient framework for testing market microstructure hypotheses as additional variables can be added to the ARMA specifications. See for example Engle and Russell (1998), Bauwens and Giot (1998) or Coppejans and Domowitz (1998). However, as explained in the introduction and using the dataset introduced in Section 2, it is possible to define the time between market events in several ways. We now introduce several possible definitions, describe their characteristics and estimate corresponding ACD and Log-ACD models.

4.2.1 No transformations

The most simple possibility is to perform no transformation on the data. Dealing with either trades or bid-ask quotes, a duration is defined as the time elapsed between two trades or two bid-ask quotes, irrespective of other information such as the price change. Using an ACD or Log-ACD model on this data, the conditional hazard gives the instantaneous trading (or quoting) intensity and Ψ_i gives the conditional expectation of the trade or quote duration. It should be stressed that a high trading intensity does not imply a corresponding price movement. For example, if the price goes back and forth between 100 \$ and 100 $\frac{1}{4}$ \$ every second,

this asset features a large trading intensity, but no price movement. It can thus be argued that these durations give relatively few information about the asset.

4.2.2 Price durations

Thinning the marked point process for the quotes with respect to a minimum change in price is one of the favourite time transformations used in the literature. See for example Bauwens and Giot (1997,1998), or Engle and Russell (1997, 1998). Price durations X_p are thus defined as the time needed to witness a given cumulative price change in the price of the asset. To avoid the bid-ask bounce exhibited by the trade process, price durations are usually defined on the mid-point of the bid-ask quote process. Let us call c_p the predefined threshold, i.e., the minimum cumulative price change that defines a duration. The set of X_p defines a new marked point process Q_p, based on the previously defined Q and P, and is characterized as $(t_{p,i}, b_{p,i}, a_{p,i})$, for $i = 1 \dots n_p$, where n_p is the total number of filtered quotes. By definition, the $t_{p,i}$ are such that the change in the mid-price on the duration $X_{p,i} = t_{p,i} - t_{p,i-1}$ is at least equal to c_p. Price durations defined in this way are very convenient for several reasons:

(i) because they are defined as the minimum amount of time for the price to increase or decrease by at least c_p, the resulting durations define an intraday volatility process. The conditional hazard is thus proportional to the instantaneous volatility, which can be characterized on an intraday basis. See below for an illustration.

(ii) as explained in Bauwens and Giot (1997) and Engle and Russell (1998), the price durations are relatively "long" with respect to the trade durations, which allows the definition of market characteristics for the trading process over the price durations for the quotes. For example, the trading intensity or average spread per transaction can be easily computed and are meaningful as they take into account a relatively large number of trades.

(iii) in a dealer's market, the bid-ask bounce can be annoying to work with, as it is a main feature of the data but gives relatively few information. Thinning the bid-ask quotes or trades allows to extract a marked point process where only meaningful price changes are retained. Moreover, as characterized in Engle and Russell (1998) for the NYSE or Biais, Hillion and Spatt (1995) for the Paris Bourse, the bid-ask quote process is often characterized by a short term transitory component which gives little information about the value of the asset. On the contrary, information events lead to movements of the bid and ask quotes in the same direction and thus move significantly the mid-point.

In Table 6, we give characteristics of price durations X_p computed using different thresholds c_p. These price durations feature a strong time-of-day effect (see

Table 6: Price durations

	$c_p = \frac{1}{8}$	$c_p = \frac{2}{8}$	$c_p = \frac{3}{8}$	$c_p = \frac{4}{8}$
Number of b-a quotes	6,728	2,193	1,026	631
Mean of $x_{p,i}$ - $X_{p,i}$	1 - 214.9	1 - 646	1 - 1287.7	1 - 1966.1
Overdisp. of $x_{p,i}$ - $X_{p,i}$	1.43 - 1.70	1.39 - 1.77	1.27 - 1.60	1.20 - 1.50
Q(10) of $x_{p,i}$	1932.22	1168.39	450.82	269.83

Data extracted from the September - November 1996 TAQ CD-ROMs for the IBM stock. The given number of bid-ask quotes is the number obtained after filtering the data (the original number of bid-ask quotes was equal to 34,321) at threshold c_p. $x_{p,i}$ is a time-of-day standardized duration, see (22), while $X_{p,i}$ are the (non standardized) filtered durations. Both are measured in seconds. The mean of $x_{p,i}$ is almost equal to 1, after the removal of the time-of-day effect. $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the $x_{p,i}$.

Figure 5) akin to the intraday seasonnality documented in Section 3. To take into account this deterministic intraday seasonality, we compute time-of-day standardized price durations, which are defined as

$$X_{p,i} = x_{p,i}\, \phi_p(t_i) \tag{19}$$

where $X_{p,i}$ is the raw filtered price duration with respect to the minimum price change c_p, $\phi_p(t_i)$ is the time-of-day effect and $x_{p,i}$ denotes the time-of-day standardized price duration. The deterministic time-of-day effect is defined as the expected price duration conditioned on time-of-day and on the day of the week (so that, for example, the time-of-day effect of Monday can be different from the time-of-day of Tuesday), where the expectation is computed by averaging the durations over thirty minutes intervals for each day of the week. Cubic splines such as used in Section 3 are then used on the thirty minutes intervals to smooth the time-of-day function.

As indicated in Table 6, increasing the minimum amount of price change needed to retain a duration decreases the number of observations and also the autocorrelation and overdispersion exhibited by the filtered durations. As could be expected, with c_p being increased, the characteristics of the filtered durations get closer to those featured by a IID Poisson distribution, i.e., no overdispersion and no autocorrelation. Nevertheless, for all given thresholds, the price durations exhibit similar characteristics (overdispersion and a strong autocorrelation). Figure 6 plots the corresponding density functions, which are also similar whatever the chosen threshold.

Exhibiting overdispersion, strong autocorrelation and a density function "similar" to the exponential distribution, the price durations are thus good candidates

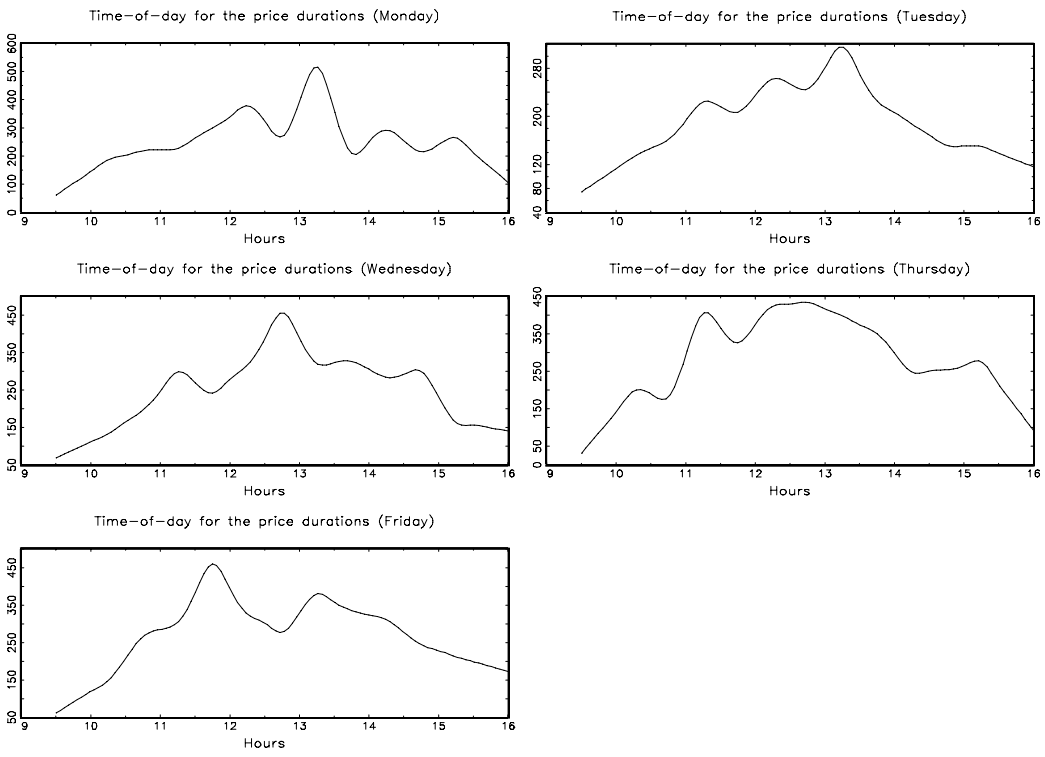

Figure 5: Intraday seasonal component for the price durations with $c_p = \frac{1}{8}$ \$. The durations are computed for the IBM stock (September 96 - November 96 or 13 weeks).

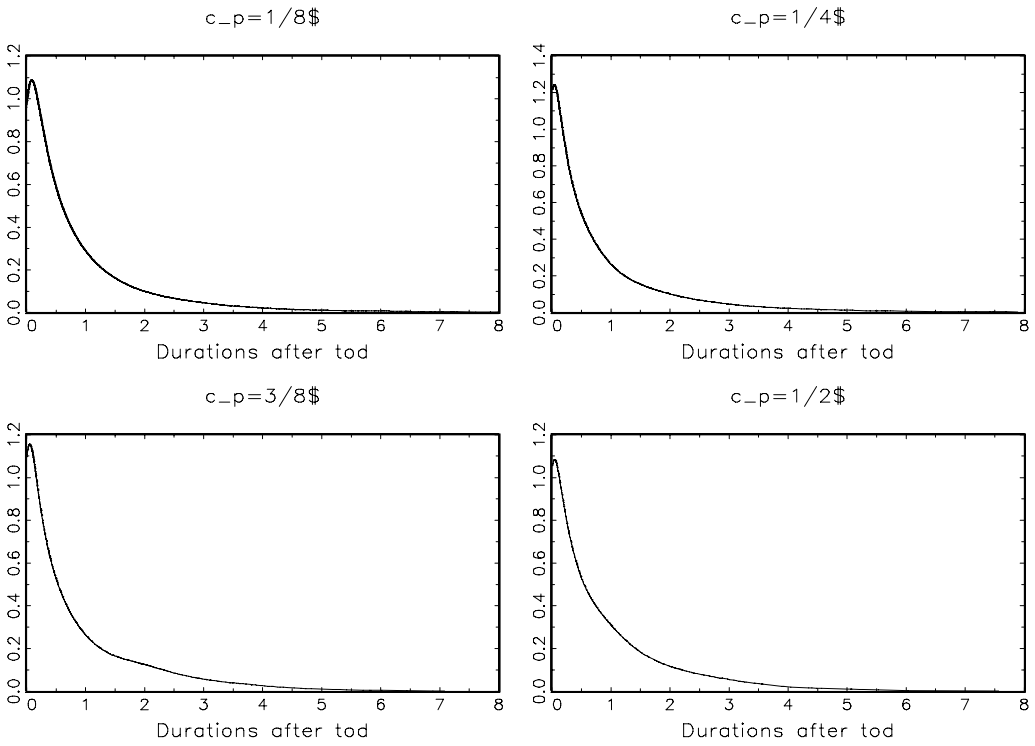

Figure 6: Density functions for the price durations (time-of-day standardized) at the specified threshold c_p. The durations are computed for the IBM stock (September 96 - November 96 or 13 weeks).

for analysis using the Weibull ACD class of models. Estimation of Log-ACD models are given in Table 7. In all cases, the Log-ACD model successfully removes the autocorrelation of the durations. The estimated coefficients are relatively stable, with β decreasing slowly as c_p is increased. As could be expected from the observed shape of the density functions, the estimated coefficient γ of the Weibull distribution is very close to one and stable for all c_p.

Table 7: ML results for the Log-ACD model
(price durations)

Coefficient	$c_p = \frac{1}{8}$	$c_p = \frac{2}{8}$	$c_p = \frac{3}{8}$	$c_p = \frac{4}{8}$
ω	-0.079 (0.007)	-0.109 (0.011)	-0.145 (0.017)	-0.172 (0.027)
α	0.077 (0.007)	0.106 (0.011)	0.141 (0.018)	0.168 (0.027)
β	0.988 (0.003)	0.985 (0.004)	0.971 (0.010)	0.955 (0.017)
γ	0.999 (0.009)	0.984 (0.016)	1.008 (0.024)	1.038 (0.031)
$Q(10)$	1932.64	1168.39	450.82	269.83
$Q(10)^*$	33.75	22.06	10.54	13.83

Estimation results for the Log-ACD model (15)-(18) applied to the price durations $x_{p,i}$ defined at threshold c_p (IBM stock, 13 weeks of intraday data). Asymptotic standard errors are given in parantheses. $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the $x_{p,i}$. $Q(10)^*$ gives the Q-statistic for the first ten autocorrelations on the estimated residuals $e_{p,i} = x_{p,i}/e^{\widehat{\psi}_i}$.

As outlined above, price durations can be used to characterize the intraday volatility. The following relationship[24] links the instantaneous intraday volatility to the conditional hazard of the price durations:

$$\sigma^2(t|I_{i-1}) = \Big(\frac{c_p}{P(t)}\Big)^2 h(x_i|I_{i-1}) \tag{20}$$

where $\sigma^2(t|I_{i-1})$ is the conditional instantaneous intraday volatility, $P(t)$ is the bid-ask quote midpoint and $h(x_i|I_{i-1})$ is the conditional hazard for the price durations defined at threshold c_p. Working with the exponential version of the Log-ACD model, $h(x_i|I_{i-1})$ reduces to $1/e^{\psi_i}$, which must be multiplied by $1/\phi_p(t_i)$ to take the deterministic intraday seasonality into account. Thus, the following relationship

$$\widehat{\sigma}^2(t|I_{i-1}) = \Big(\frac{c_p}{P(t)}\Big)^2 \frac{1}{e^{\widehat{\psi}_i}\phi_p(t_i)} \tag{21}$$

[24]See Engle and Russell (1998) for a formal proof.

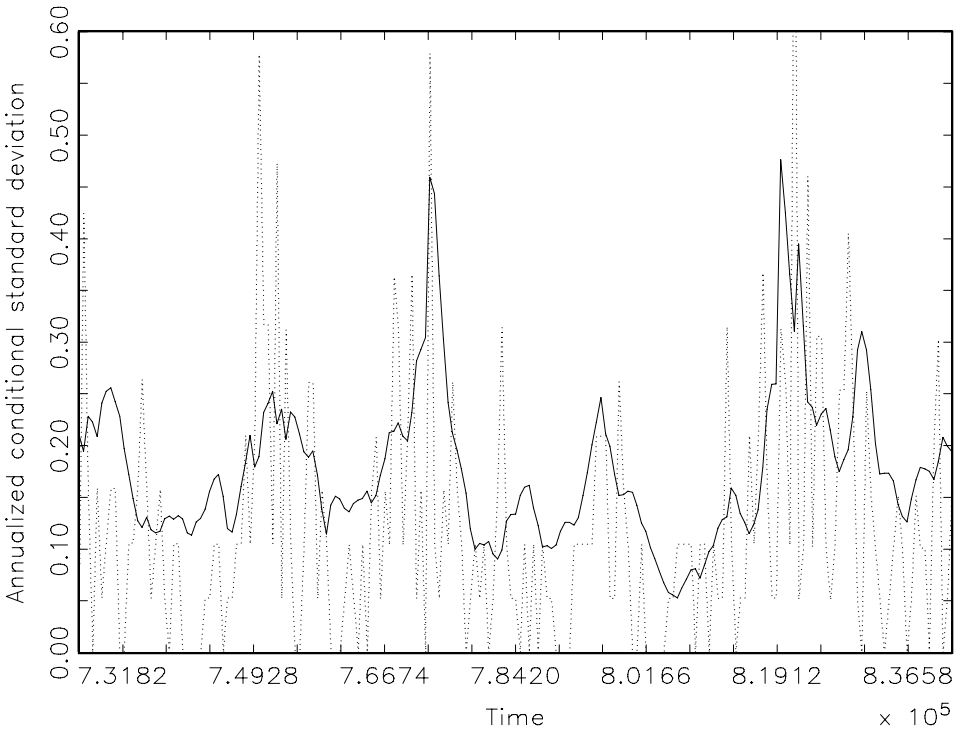

Figure 7: Annualized conditional standard deviation as forecasted by the EGARCH model (continuous line) and observed absolute returns. The returns are defined on the mid-point of the bid-ask quotes of the IBM stock (h=10 minutes). The model is estimated for the September 96 - October 96 period and the results are plotted for the second week of September.

provides a direct estimation of the intraday volatility once the Log-ACD model has been estimated. With $c_p = \frac{1}{8}$ \$ and using the correponding estimate of $e^{\widehat{\psi}_i}$ as given by the estimation of the Log-ACD model, Figure 8 plots the annualized equivalent[25] of $\widehat{\sigma}(t|I_{i-1})$ for the second week of September. The pattern of intraday volatility for the second week of September was previously computed in Section 3 using an EGARCH model. To allow easy comparison with these alternative results, Figure 7 reproduces Figure 3, but with the intraday volatility indexed by the number of seconds (the same scale is used in Figure 8). As indicated by Figure 8, the ACD-based estimation of the conditional volatility is an alternative to the ARCH modelling performed on equidistantly sampled data.

[25]It is computed as $\sqrt{\widehat{\sigma}(t|I_{i-1})250(16-9.5)3600}$.

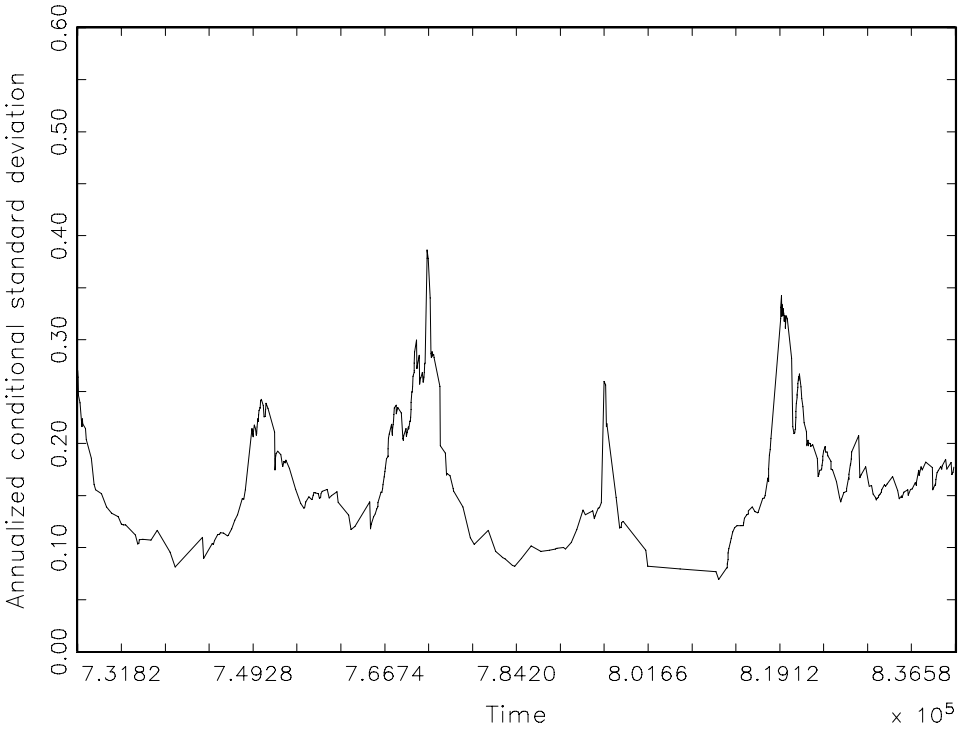

Figure 8: Annualized conditional standard deviation as forecasted by the Log-ACD model estimated on the price durations with $c_p = \frac{1}{8}$ \$. The model is estimated for the September 96 - October 96 period and the results are plotted for the second week of September.

4.2.3 Volume durations

The previously defined price durations use information given by the bid-ask prices associated to the observed durations in order to define new durations related to the intraday volatility process. As we focus on the marked point process for the bid-ask quotes Q, another possibility is to use the information given by the traded volume. Indeed, market participants could be interested by the time needed to trade a given volume c_v of shares. Thinning the quote process such that the retained durations are characterized by a total traded volume at least equal to c_v defines a new set of durations $X_{v,i}$, called volume durations[26].

Volume durations have an immediate appeal for characterizing the liquidity of a stock, as a short X_v imply that a given volume can be traded quickly. The set of X_v defines a new marked point process Q_v, based on the previously defined Q and P, which is characterized as $(t_{v,i}, b_{v,i}, a_{v,i})$, for $i = 1 \dots n_v$, where n_v is the total number of filtered quotes. By definition, the $t_{v,i}$ are such that the traded volume on duration $X_{v,i} = t_{v,i} - t_{v,i-1}$ is at least equal to c_v. The change in price (either the bid, the ask or the mid-point) over $X_{v,i}$ can be interpreted as the price response to a traded volume equal to at least c_v shares. A liquid stock would be characterized by short $X_{v,i}$ with small price changes over these $X_{v,i}$, i.e., that it is possible to trade a large amount of shares in a small amount of time without having a large impact on the price[27].

Using the IBM database for the trades and quotes, we compute the volume durations $X_{v,i}$ for several thresholds c_v. As explained above, these new durations are defined for the bid-ask quote process with respect to the trade process which gives information about the traded volume. As indicated in Figure 9, volume durations are characterized by a strong intraday effect, similar to the one documented for "normal" and price durations. This leads us to define time-of-day standardized durations $x_{v,i}$ as

$$X_{v,i} = x_{v,i}\, \phi_v(t_i) \tag{22}$$

where $X_{v,i}$ is the raw filtered volume duration with respect to the minimum traded volume c_v $\phi_v(t_i)$ is the time-of-day effect and $x_{v,i}$ denotes the time-of-day standardized volume duration. The deterministic time-of-day effect is defined as the expected volume duration conditioned on time-of-day and on the day of the week, with cubic splines used to smooth the function.

Descriptive characteristics about these volume durations are given in Table 8 and the corresponding density functions are given in Figure 10. As indicated in Table 8, volume durations are characterized by a very large autocorrelation (this feature was already observed when dealing with "normal" and price durations)

[26] Gouriéroux, Jasiak and Le Fol (1996) introduce volume durations for the trade process.

[27] Kyle (1985) introduced three notions of liquidity: tightness (bid/ask spread), depth (amount of one sided volume that can be absorbed by the market without causing a revision of the bid/ask prices) and resiliency (speed of return to the equilibrium).

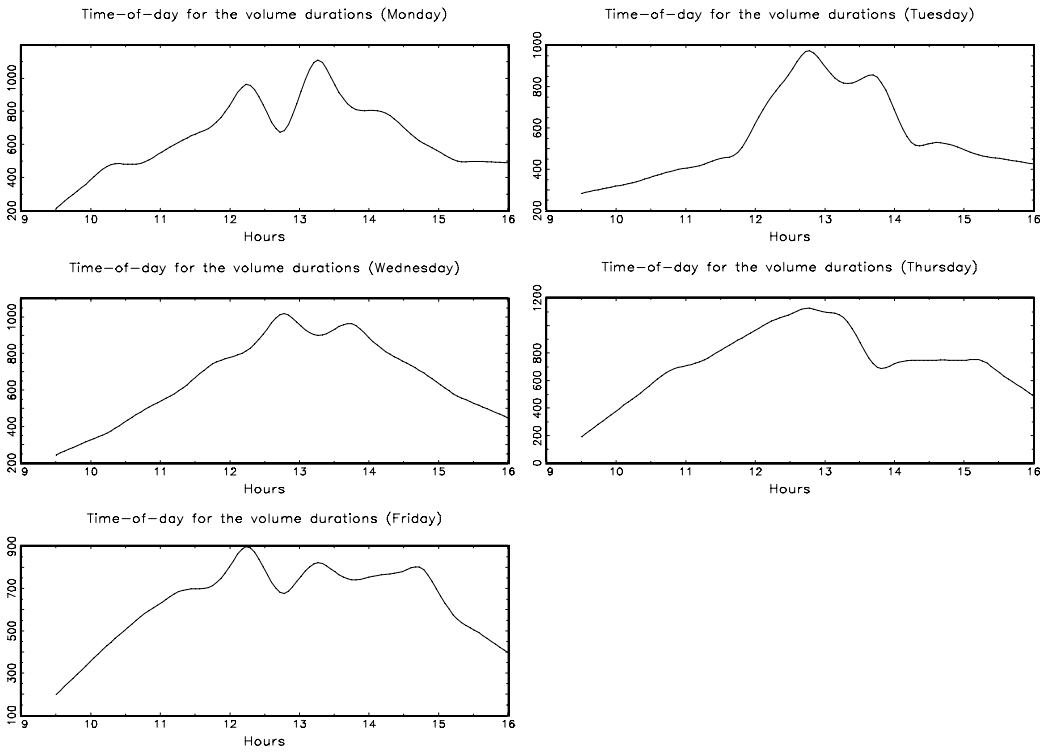

Figure 9: Intraday seasonal component for the volume durations with $c_v = 50,000$ shares. The durations are computed for the IBM stock (September 96 - November 96 or 13 weeks).

Table 8: Volume durations

	$c_v = 25000$	$c_v = 50000$	$c_v = 100000$	$c_v = 200000$
Number of b-a quotes	4,305	2,442	1,302	658
Mean of $x_{v,i}$ - $X_{v,i}$	1 - 336.5	1 - 589	1 - 1091.2	1 - 2089.4
Overdisp. of $x_{v,i}$ - $X_{v,i}$	0.80 - 0.93	0.71 - 0.85	0.63 - 0.78	0.56 - 0.71
Q(10) of $x_{v,i}$	3476.48	3378.34	2150.2	1213.55

Data extracted from the September - November 1996 TAQ CD-ROMs for the IBM stock. The given number of bid-ask quotes is the number obtained after filtering the data (the original number of bid-ask quotes was equal to 34,321) at threshold c_v. $x_{v,i}$ is a time-of-day standardized duration, see (22), while $X_{v,i}$ are the (non standardized) filtered durations. Both are measured in seconds. The mean of $x_{v,i}$ is almost equal to 1, after the removal of the time-of-day effect. $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the $x_{v,i}$.

and also exhibit underdispersion (this is in sharp contrast with "normal" and price durations, which are characterized by overdispersion).

The density functions given in Figure 10 are quite different from the familiar density functions observed for price durations. While the latter were characterized by a density function similar to the exponential distribution for all c_p, density functions for volume durations become hump-shaped as c_v is increased, with most of the mass around 1. Estimation of Log-ACD models for volume durations $X_{v,i}$ computed for several thresholds c_v are given in Table 9. As could be expected from the general shape of the density functions, the coefficient γ is much greater than one, which is needed to exhibit the hump-shaped function. The results given in Table 9 indicate that the Log-ACD model is quite successful in removing the autocorrelation exhibited by the volume durations. For all c_v, coefficient β is close to one and decreases slightly when c_v is increased.

In the recent literature (Bauwens and Giot, 1997, 1998; Engle and Russell, 1997, 1998), the ACD and Log-ACD models have been applied to data featuring overdispersion. As indicated by Engle and Russell (1998), this is one of the main motivation for the use of the ACD class of models. As volume durations feature underdispersion, it is thus interesting to check if the Log-ACD model can accomodate this characteristic. For all four estimated specifications, we conduct Monte Carlo simulations and generate 200 chains of 10,000 observations. The average underdispersions are given in Table 10 and indicate that the underdispersion of the simulated data is close to the one featured by the data used to estimate the model (see Table 8). Thus, according to the evidence given in Table 6-9, the Log-ACD model can successfully deal with durations that are correlated and that feature either under or overdispersion.

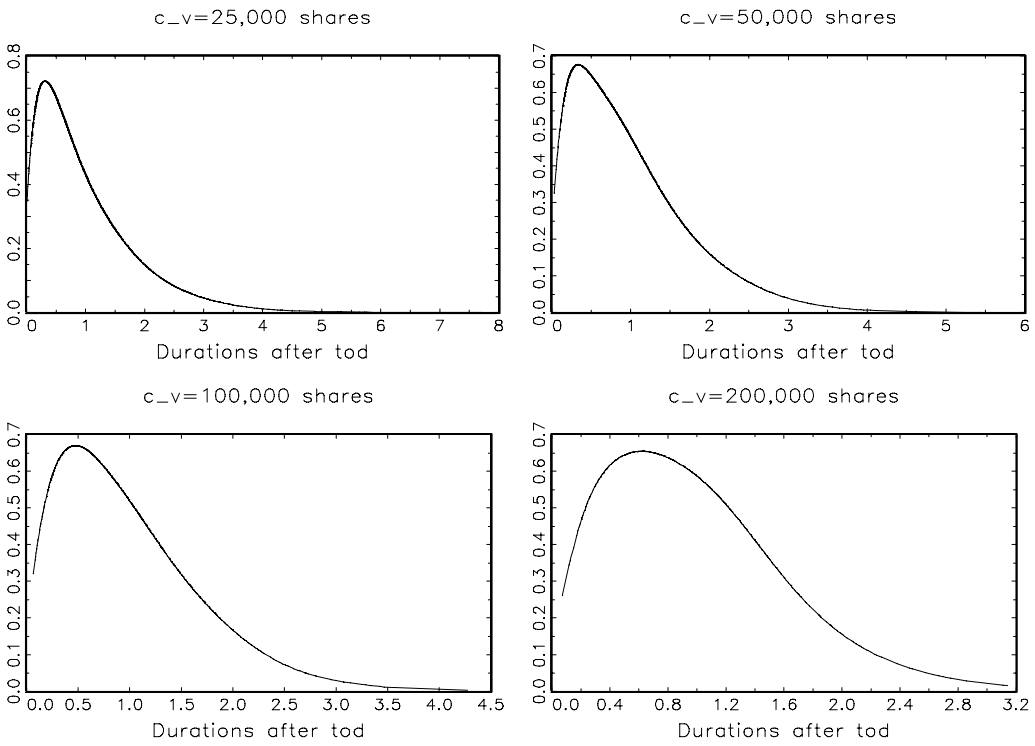

Figure 10: Density functions for the volume durations (time-of-day standardized) at the specified threshold c_v. The durations are computed for the IBM stock (September 96 - November 96 or 13 weeks).

Table 9: ML results for the Log-ACD model
(volume durations)

Coefficient	$c_v = 25000$	$c_v = 50000$	$c_v = 100000$	$c_v = 200000$
ω	-0.134 (0.010)	-0.188 (0.014)	-0.247 (0.020)	-0.331 (0.032)
α	0.148 (0.011)	0.211 (0.016)	0.277 (0.023)	0.369 (0.036)
β	0.987 (0.003)	0.981 (0.005)	0.973 (0.008)	0.945 (0.015)
γ	1.644 (0.019)	1.930 (0.029)	2.222 (0.046)	2.488 (0.071)
$Q(10)$	3789.58	3379.20	2150.20	1213.55
$Q(10)^*$	37.72	24.20	13.35	14.55

Estimation results for the Log-ACD model (15)-(18) applied to the volume durations $x_{v,i}$ defined at threshold c_v (IBM stock, 13 weeks of intraday data). Asymptotic standard errors are given in parantheses. $Q(10)$ denotes the Ljung-Box Q-statistic for the first ten autocorrelations on the $x_{v,i}$. $Q(10)^*$ gives the Q-statistic for the first ten autocorrelations on the estimated residuals $e_{v,i} = x_{v,i}/e^{\widehat{\psi}_i}$.

5 Conclusion

In this paper, we have presented two different ways of dealing with intraday data for a stock traded at the New York Stock Exchange (NYSE). Using the Trade and Quote database made available by the NYSE, we characterized the trade and quote datasets on which the econometric models are applied. First, we focused on regularly sampled data, which allows for the estimation of (intraday) GARCH and EGARCH models on the mid-point of the bid-ask quote process. As indicated previously by Andersen and Bollerslev (1997), taking into account the intraday seasonality is an important feature of the analysis. Combining the deterministic intraday pattern for the volatility with the forecasted stochastic component by the EGARCH model, the precise pattern of intraday volatility can be obtained. Augmenting the dataset with information given by the trade process (such as the traded volume, number of trades and average volume per trade), EGARCH models that include these variables in the specification of the log-variance are estimated. Including these variables has a important effect on the estimated coefficients of the EGARCH model, which supports the fact that the traded volume and number of trades are among the main determinants of the intraday volatility process[28].

Secondly, drawing on the recent literature about high-frequency duration models, we deal with the irregularly sampled marked point process for the bid-ask

[28]Lamoureux and Lastrapes (1992) and Jones, Kaul and Lipson (1994) conduct similar analysis on daily data and reach similar conclusions.

Table 10: Simulated data
(volume durations)

Coefficients	$c_v = 25000$	$c_v = 50000$	$c_v = 100000$	$c_v = 200000$
ω	-0.134	-0.188	-0.247	-0.331
α	0.148	0.211	0.277	0.369
β	0.987	0.981	0.973	0.945
γ	1.644	1.930	2.222	2.488
Overdisp.	0.91	0.84	0.77	0.66

Monte Carlo simulations. For the four sets of parameters estimated in Table 9, the overdispersion is computed by simulating 200 chains of 10,000 draws from the Weibull distribution.

quotes. We define two time transformations that thin the original datasets so that price durations (durations that lead to a cumulative price change equal to a given threshold) and volume durations (durations that lead to a cumulative traded volume equal to a given threshold) are defined. Price (volume) durations are closely linked to the intraday volatility (liquidity) process. While the price durations feature a strong autocorrelation and overdispersion, the volume durations exhibit autocorrelation and underdispersion. In both cases, the Log-ACD model is successfully applied to the newly defined datasets. One of the interesting by-products of price durations and ACD type models is that they allow an easy computation of intraday volatility, which provides an alternative to the ARCH class of models.

References

[1] Andersen, T.G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115-158.

[2] Bauwens, L. and 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.*

[3] Bauwens, L. and Giot, P. (1998). Asymmetric ACD models: introducing price information in ACD models with a two state transition model. *CORE Discussion Paper 9844.*

[4] Bauwens, L. and Veredas, D. (1999). Stochastic conditional duration models. *Mimeo.*

[5] Biais, B., Hillion, P. and Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance* 50, 1655-1689.

[6] Biais, B., Foucault, T. et Hillion, P. (1997). *Microstructure des marchés financiers: institutions, modèles et test empiriques*. Paris: Presses Universitaires de France.

[7] Bisière, C. and Kamionka, T. (1998). Timing of orders, orders aggressiveness and the order book at the Paris Bourse. *Mimeo*. GREMAQ, Toulouse.

[8] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[9] Bollerslev, T. and Domowitz, I. (1993). Trading patterns and prices in the interbank foreign exchange market. *Journal of Finance* 48, 1421-1443.

[10] Brock, W.A. and Kleidon, A.W. (1992). Periodic market closure and trading volume: a model of intraday bids and asks. *Journal of Economic Dynamics and Control* 16, 451-489.

[11] Coppejans, M. and Domowitz, I. (1998). Stock and flow information as inputs to limit order book trading activity. *Mimeo*.

[12] Darolles, S. , Gouriéroux, C. and Le Fol, G. (1998). Intraday transaction price dynamics. *Mimeo*. CREST, Paris.

[13] Easley, D. and O'Hara, M. (1992). Time and the process of security price adjustment. *The Journal of Finance* 19, 69-90.

[14] Easley, D., Kiefer, N.M. and O'Hara, M. (1997). The information content of the trading process. *The Journal of Empirical Finance* 159-186.

[15] Ederington, L.H. and Lee, J.H. (1993). How markets process information: news releases and volatility. *Journal of Finance* 48, 1161-1191.

[16] Engle, R. and Lunde, A. (1998). Trades and quotes: a bivariate point process. *Discussion Paper 98-07*. University of California, San Diego.

[17] Engle, R. and Russell, J. (1997). Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance* 4, 187-212.

[18] Engle, R. and Russell, J. (1998). Autoregressive conditional duration; a new model for irregularly spaced transaction data. *Econometrica* 66, 1127-1162.

[19] Ghysels, E. and Jasiak, J. (1997). GARCH for irregularly spaced financial data: the ACD-GARCH model. *Discussion Paper 97s-06*. CIRANO, Montréal.

[20] Glosten, L.R. and Harris, L.E. (1988). Estimating the components of the bid-ask spread. *Journal of Financial Economics* 21, 123-142.

[21] Glosten, L. and Milgrom, P. (1985). Bid, ask, and the transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 13, 71-100.

[22] Goodhart, C.A.E. and O'Hara, M. (1997). High frequency data in financial markets: issues and applications. *The Journal of Empirical Finance* 73-114.

[23] Gouriéroux, C.J., Jasiak, J. and Le Fol, G. (1996). Intraday market activity. *CREST Working Paper 9633.*

[24] Grammig, J., Hujer, R., Kokot, S. and Maurer, K.O. (1998). Modeling the Deutsche Telekom IPO using a new ACD specification. *Discussion Paper 55/1998.* National Research Center on Quantification and Simulation of Economic Processes, Humboldt University Berlin.

[25] Grammig, J. and Wellner, M. (1999). Modeling the interdependence of volatility and inter-transaction duration processes. *Discussion Paper 21/1999.* Humboldt University Berlin.

[26] Guillaume, D.M., Dacorogna, M.M., Davé, R.R., Muller, U.A., Olsen, R.B. and Pictet, O.V. (1995). From the bird's eye to the microscope: a survey of new stilized facts of the intra- daily foreign exchange markets. *Olsen Preprint*

[27] Guillaume, D.M., Dacorogna, M.M. and Pictet, O.V. (1997). On the intra-daily performance of GARCH processes. *Olsen Preprint*

[28] Gwilym, O.A., Buckle, M. and Thomas, S. (1997). The intraday behavior of bid-ask spreads, returns, and volatility for FTSE-100 Stock Index Options. *Journal of Derivatives* 4, 20-32.

[29] Hafner, C. (1996). Estimating high frequency foreign exchange rate volatility with nonparametric ARCH models. *Mimeo (preliminary version).*

[30] Hasbrouck, J. (1991). Measuring the information content of stock trades. *Journal of Finance* 46, 179-208.

[31] Jones, C.M., Kaul, G. and Lipson, M.L. (1994). Transactions, volume, and volatility. *Review of Financial Studies* 7, 631-651.

[32] Kyle, A.S. (1985). Continuous auctions and insider trading. *Econometrica* 53, 1315-1365.

[33] Lamoureux, C.G. and Lastrapes, W.D. (1990). Heteroskedasticity in stock return data: volume versus GARCH effects. *Journal of Finance* 45, 220-229.

[34] Lee, C.M. and Ready, M.J. (1991). Inferring trade direction from intraday data. *Journal of Finance* 46, 733-746.

[35] Le Fol, G. and Mercier, L. (1998). Time deformation: definition and comparisons. *Journal of Computational Intelligence in Finance* September/October 1998, 19-33.

[36] Maillet, B. and Michel, T. (1998). Volume-time scale and intra-day returns density. *Mimeo*.

[37] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 349-370.

[38] O'Hara, M. (1995). *Market microstructure theory*. Oxford: Basil Blackwell.