Econometric Analysis of Intra-Daily Trading Activity on the Tokyo Stock Exchange

Luc Bauwens

We analyze statistically inter-trade durations of four stocks listed on the Tokyo Stock Exchange in 2003. We find that these data display the usual stylized facts (intra-daily seasonality, clustering, and overdispersion) found for similar data of the New York Stock Exchange, but with some differences. We also estimate autoregressive conditional duration models for fitting the durations. We find that, as with comparable data of the NYSE, some models fit in a satisfactory way the dynamic properties of the durations, but do not always fit well the conditional distribution of the data.

Keywords: Autoregressive conditional duration; Trade durations; High-frequency data; Tokyo Stock Exchange
JEL Classification: C10, C41, G10

CORE and Department of Economics, Université catholique de Louvain (E-mail: bauwens@core.ucl.ac.be)

This paper was written during my stay as a visiting scholar at the Institute for Monetary and Economic Studies (IMES) of the Bank of Japan (BOJ). I thank very much the BOJ for providing the financial and logistic support to conduct this research during my stay. I am grateful to Hironobu Tozaka for preparing the data and help in understanding them, and to Helena Beltran-Lopez and Andreas Heinen for interesting discussions about the data. While remaining responsible for any error in the paper, I thank the Financial Engineering Team of IMES and an anonymous referee for many suggestions that helped me to improve the paper.

I. Introduction

In recent years, there has been a lot of research, both theoretical and empirical, on financial market microstructure. Several theoretical contributions emphasize that the waiting times between intra-day market events such as trades, quote updates, price changes, and order arrivals play a key role for understanding the processing of private and public information in financial markets.[1] On the empirical side, the accessibility of high-frequency data at a micro level, which ideally includes real-time recordings of trades, order arrivals, and quote updates, as well as the corresponding prices, volumes, and time stamps, opened new perspectives for the empirical analysis of the market microstructure of financial markets.

In econometrics, one area of research that has developed directly in connection with the empirical aspects just mentioned is duration analysis in a dynamic framework. An econometric model of serially correlated event arrival times was provided by Engle and Russell (1998), who introduced the autoregressive conditional duration (ACD) model. The ACD approach combines elements from transition analysis (Lancaster [1990]) and Engle's (1982) autoregressive conditional heteroskedasticity (ARCH) model. Indeed, the motivation behind the ACD and the ARCH models appears similar: financial market events, such as trades and quote changes, occur in clusters. Following the contribution of Engle and Russell (1998), several modifications of the basic models have been put forward. Bauwens and Giot (2000) introduced a logarithmic version of the ACD model, which implies a nonlinear relation between the duration and its lags. As an alternative to the exponential and Weibull distributions used in Engle and Russell's (1998) seminal paper, Lunde (2000) and Grammig and Maurer (2000) considered ACD specifications based on the generalized gamma (GG) and the Burr distribution (both nest Weibull and exponential as special cases). Zhang, Russell, and Tsay (2001) advocated the threshold ACD model, formulated in the spirit of threshold autoregressive models, to capture a possible nonlinear relation between the duration and predetermined variables and to account for regime switches. Ghysels, Gouriéroux, and Jasiak (2004) developed the stochastic volatility duration (SVD) model, a dynamic two-factor model, which is designed to account both for mean and variance dynamics in financial duration processes. Another (single) factor model, the stochastic conditional duration (SCD) model, was put forth by Bauwens and Veredas (2003).

Bauwens *et al.* (2004) compare and evaluate most of these duration models on trade durations pertaining to stocks traded on the New York Stock Exchange (NYSE). In this paper, first we analyze statistically trade durations of four stocks listed on the Tokyo Stock Exchange (TSE) in 2003. We find that these data display the stylized facts of inverted-U shape intra-daily pattern, clustering and overdispersion, found for similar data of the NYSE. We also find that the trade durations of the stocks we study feature less overdispersion than for NYSE stocks, and even in some cases the durations are slightly underdispersed. However, it is quite likely that

1. See O'Hara (1995) for many details and references, and Goodhart and O'Hara (1997) and Madhavan (2000) for more recent surveys.

the duration dispersion is artificially underestimated due to the way the data are recorded, since actually small durations are not precisely measured.

Second, we estimate ACD models for the durations. We find that, as with comparable data for the NYSE, some models fit in a satisfactory way the dynamic properties of the durations, but do not always fit well the conditional distribution of the data.

The structure of the paper is as follows: in Section II, we describe the features of the trading mechanism of the TSE that are required to understand the data that we model; Section III serves to describe the data; in Section IV we review duration models and report the empirical results, and Section V offers our conclusions.

II. The TSE Trading System

This section serves to describe briefly the TSE trading system, in particular, the features that help to understand the data we model. As these data consist of durations between trades executed on the market for a given stock, it is important to understand how the trades are generated and how the data are collected. For a more comprehensive description of the trading method on the TSE, we refer to Tokyo Stock Exchange (2003).

In 2003, which is the year for which we have data, the TSE was functioning as an order-driven market. We refer to Bauwens and Giot (2001, chapter 1) for a review of market types, in particular the distinction between price-driven and order-driven markets. In an order-driven market, trading participants, that is, securities companies licensed by the exchange to trade on the market, may enter two types of orders: limit orders and market orders. Each type of order may be a buy order or a sell order. In fact, other types of orders exist on the TSE, which are limit or market orders only effective under certain conditions, but they are not relevant for our analysis.

Each limit order specifies a quantity for sale at a given minimum price (called the ask or offer price) or a quantity to buy at a given maximum price (called the bid price). At any time, the entire set of limit orders constitutes the order book. Normally, the best (i.e., lowest) ask price is strictly larger than the best (i.e., highest) bid price, in which case no exchange is possible. Traders who issued the limit orders are waiting for other traders to match their limit orders with market orders or limit orders. A trader who wants to buy or sell immediately issues a market order for a given quantity, meaning that he or she is willing to buy or sell no more than the specified volume, at the best available price. Hence, limit orders provide liquidity, while market orders consume it.

The top panel of Table 1 illustrates an order book with three limit-sell orders, with the lowest ask price at 100 and the highest price at 104, and two limit-buy orders at prices of 98 and 96. The middle panel of the same table shows the state of the order book after the execution of a market buy order of 1,000 shares. This order has been crossed with the lowest limit-sell order, hence the transaction price is 100. The bottom panel of the table shows the state of the book after the execution of a subsequent market sell order of 1,500 shares. This order exhausts the highest limit-buy order (750 units at 98) and consumes 750 shares of the next-best limit-buy order (at the price of 96). In this case, the average price per share is 97. Another

Table 1 Order Book

Sell-side volume	Unit price	Buy-side volume
Start state		
1,000	104	
1,800	102	
4,000	100	
	98	750
	96	2,000
State after a market buy order of 1,000 shares		
1,000	104	
1,800	102	
3,000	100	
	98	750
	96	2,000
Next state after a market sell order of 1,500 shares		
1,000	104	
1,800	102	
3,000	100	
	96	1,250

possibility is that only 750 shares are exchanged at the price of 98, while the remaining 750 shares are entered in the order book as a limit-sell order at the executed price of 98 (it is also possible to cancel automatically the non-executed part of the order). In this example, two trades have occurred, corresponding to two executed market orders.

The trading system described above functions during the trading sessions of the TSE (it is named the *Zaraba* method). There are two trading sessions per weekday: the morning session, which starts at 9:00 and ends at 11:00, and the afternoon session, from 12:30 until 15:00. Market and limit orders are submitted during these sessions, though limit orders may be entered also before the start of the sessions: from 8:00 until 9:00 and from 12:05 until 12:30. The opening price of each trading session is the result of an opening auction, and the closing price is the result of a closing auction. These auctions correspond to a Walrasian-type auction (named the *Itayose* method), in that demands and offers are accumulated (resulting from the limit orders), and a price which clears the market as much as possible is fixed. Of course, this usually results in transactions. The system provides a "clean" order book at the beginning of a session, that is, there is no bid at a price higher than the best ask price and no ask at a price lower than the best bid price.

III. Data Description

In this section, we describe what raw data we have used (Section III.A), what transformations we have applied to these data (Section III.B), and what are the main statistical features of the data from a descriptive viewpoint (Section III.C).

A. Raw Data

We use intra-daily data for four major stocks listed on the TSE: Nippon Steel Corp. (NPS), Sony Corp. (SON), Tokyo Electric Power Co., Inc. (TKE), and Toyota Motor Corp. (TOY). The data are from Bloomberg. The sampling period corresponds to the months from March until July 2003. For each stock and each trading day, the raw data are as shown in columns 2 to 5 in Table 2: records 1 to 13 correspond to the beginning, and records 14 to 31 to the end of the morning session. A record with the label "bid price" ("ask price") presumably provides the best bid (ask) price and the corresponding buy (sell) volume available at this price in the order book; this volume may of course

Table 2 Raw Data Example

Record	Time	Label	Price	Volume	Duration
1	08:59:56	Bid price	164.000	2,624,000	
2	08:59:56	Ask price	165.000	3,497,000	
3	09:00:10	Bid price	164.000	2,902,000	
4	09:00:10	Ask price	165.000	3,545,000	
5	09:00:26	Last trade	164.000	2,216,000	—
6	09:00:26	Bid price	164.000	686,000	
7	09:00:26	Ask price	165.000	1,338,000	
8	09:00:30	Last trade	165.000	30,000	4
9	09:00:30	Bid price	164.000	686,000	
10	09:00:30	Ask price	165.000	1,308,000	
11	09:00:34	Last trade	164.000	10,000	4
12	09:00:34	Bid price	164.000	676,000	
13	09:00:34	Ask price	165.000	1,308,000	
...					
14	10:59:32	Bid price	162.000	228,000	
15	10:59:32	Ask price	163.000	1,382,000	
16	10:59:48	Last trade	162.000	1,000	—
17	10:59:48	Bid price	162.000	227,000	
18	10:59:48	Ask price	163.000	1,382,000	
19	10:59:50	Bid price	162.000	209,000	
20	10:59:50	Ask price	163.000	1,382,000	
21	11:00:00	Bid price	162.000	199,000	
22	11:00:00	Ask price	163.000	1,382,000	
23	11:00:00	Last trade	162.000	10,000	12
24	11:00:02	Bid price	162.000	202,000	
25	11:00:02	Ask price	163.000	1,382,000	
26	11:00:06	Last trade	162.000	24,000	6
27	11:00:06	Bid price	162.000	180,000	
28	11:00:06	Ask price	163.000	1,382,000	
29	11:01:24	Bid price	0.000		
30	11:01:24	Ask price	0.000		
31	11:02:06	Last trade	159.000		

Note: Data for NPS stock on March 3, 2003. Price in yen. Volume in number of shares. Duration in seconds.

Source: Bloomberg.

be the sum of bids (asks) of different traders. A record with the label "last trade" corresponds to a transaction presumably due to a market order. For example, the trade reported in record 8 is done at the ask price available at 9:00:26 (record 7), and it results from executing a market buy order for 30,000 units of volume. One can see that the new ask volume is 1,308,000 (record 10), down from 1,338,000, the volume on the ask side available before the market order was executed. Notice that it is not always possible from the data to know exactly what has happened, since the data do not actually provide full information about the order book. For example, orders may be cancelled, and new orders may arrive in the system, such that the change of the bid (or ask) volume is not necessarily equal to the amount reported in the last trade.

A few special features of the data appear. The first one is that some mistakes appear (see records 29–31), but they are obviously linked to the closing. Such records can be eliminated without harm. The second one is that the order of the records is seemingly not always as it should be: usually the bid and ask price records reflecting the new situation after a trade appear just after the "last trade" record, with the same time stamp (see, e.g., records 8–10, 11–13, and others), but in some cases they appear just before it (e.g., records 21–23). The third feature is the fact that all time stamps end with an even number of seconds. This is quite unusual and implies that any duration between two events is an even number if measured in seconds. Our understanding is that the TSE system sends the data whenever a new trade or change of the best bid/ask situation occurs, with time stamps to the minute, and that Bloomberg affixes the seconds corresponding approximately to moments when it receives the data. It seems that in this process the seconds in the time stamps are somewhat rounded to even numbers. No explanation for this is available from the data vendor. Therefore, we consider for the analysis of the data that the precision of time recording is of two seconds, rather than one second, as in the Trades and Quotes (TAQ) database of the NYSE.[2] Finally, there are zero durations between some trades, which is not surprising given the data precision. Two orders executed almost at the same time (within two seconds) have an identical time stamp. We consider that the trades have been executed in the order in which they are recorded, and we choose to assign a duration of one second between such "simultaneous" trades. This is more correct than just discarding the zero durations, as this would truncate the distributions of the durations at two seconds.

As we are interested in the analysis of durations between trades, we select all correct records with the "last trade" label, provided they are time-stamped in the range 9:00:00–11:00:59 in the morning, and 12:30:00–15:00:59 in the afternoon. We keep the trades stamped to the first minute after the official closing times because we suspect there is some delay in reporting the trades, that is, we choose to consider them as normal trades. From these records, we compute the inter-trade durations of each session. The first duration of a session is the one between the first two trades after the opening, and the last one corresponds to the difference between the time stamps of the last two trades (see the last column of Table 2). The duration between the last trade of a session and the first trade of the next session is obviously not used, since it would be artificially long. Finally, we divide all durations by two, that is,

2. See Bauwens and Giot (2001, chapter 2) for a description of the TAQ database.

the unit for measuring durations is two seconds. In this way, we avoid artificially nonexisting odd durations.

For further use, we denote by t_i the value of the time point of the i-th trade of a stock, and by X_i the corresponding "raw" duration, that is, $X_i = t_i - t_{i-1}$.

B. Intra-Daily Seasonality of Durations

The next step consists of adjusting the raw durations for their intra-daily seasonal pattern. Indeed, it is well known that activity on stock markets is subject to variations linked to the time of the day. This is due to the institutional features of the exchanges, such as the opening and closing times, and the habits of traders (in particular, the lunchtime effect; however, this is not relevant to the TSE since there is a break from 11:00 to 12:30). Usually trading activity is the most intense, hence inter-trade durations are the shortest (on average), at the beginning and at the end of the day. At the start of the day, trading is very active as the opening of the market prompts traders to take positions such that the information brought about by news events that occurred before the opening (macroeconomic news, or news released by companies after the previous market close) is included in the prices of assets. High trading activity at the end of the day is partly justified by the fact that traders often wish to close their positions before the end of the trading session. Given that a midday break is a feature of the TSE, it is of interest to know if the general pattern of intra-daily seasonality of durations on this market is fundamentally different from that of exchanges which do not organize a break (the inverted-U shape alluded to above).[3]

The intra-daily seasonal pattern may be defined as the expectation of the variable of interest (trade duration) conditioned on the time of day. We estimate the expectation by averaging the observed durations over 30-minute intervals for each day of the month. This is equivalent to assuming that the trading day is divided into intervals of 30 minutes (four in the morning, five in the afternoon), and that each expectation is constant in the 30-minute interval. The last hypothesis is obviously too coarse in practice, as each expectation may change throughout the trading day. Assuming that such changes happen gradually, cubic splines are then used on the 30-minute intervals to smooth the time-of-day functions. In doing so, the morning and the afternoon sessions are treated separately.

Figures 1 and 2 display the time-of-day functions of the durations of the four stocks in our database, for the months of March until July 2003. For SON and TOY, the functions are at about half of the level of the other stocks (TKE and NPS), corresponding to a higher activity level (and therefore smaller durations). For the morning session, the time-of-day function has in most cases an inverted-U pattern with the starting point lower than the end point (it is higher in a few cases). For the afternoon session, the function is in most cases decreasing, sometimes with a flat or even slightly increasing section in the beginning. Note that the starting point in the afternoon is in all cases larger than the end point in the morning. Over a complete day, we recover essentially the inverted-U shape found in other exchanges, albeit with the change of level at 12:30 due to the midday break.

3. See Bauwens and Giot (2001, chapter 2) for illustrations in the case of the NYSE.

Figure 1 Time-of-Day Functions (In Two Seconds), All Months, NPS and TOY

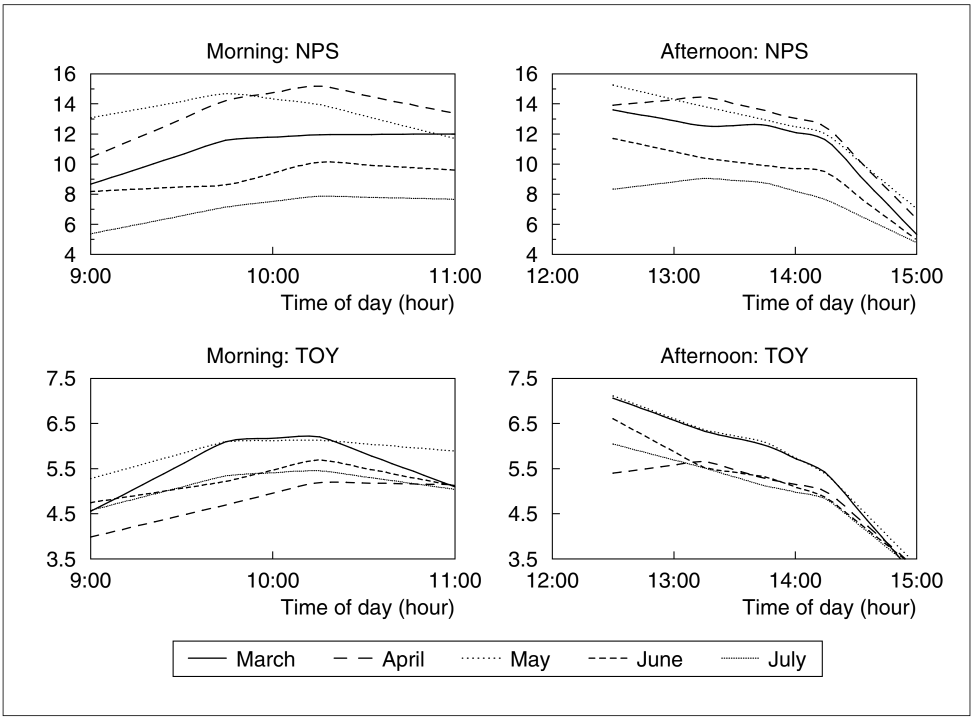
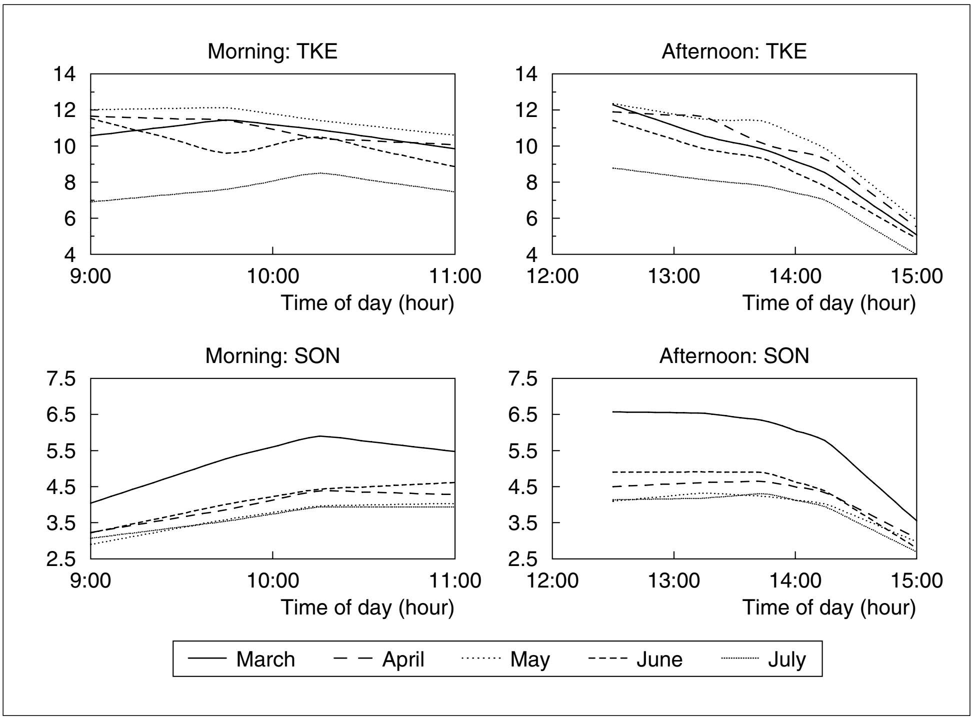


Figure 2 Time-of-Day Functions (In Two Seconds), All Months, TKE and SON



Over the period March–July 2003, there is seemingly no general trend in the level of activity. For NPS and TKE, the level of the curves for June and especially July is lower than for the previous months, which may indicate a positive trend in activity. For TOY and SON, there is no such clear pattern. We have no information on why there may be a positive trend in activity of NPS and TKE, and not in the other stocks. Explaining these differences could be of interest, but should certainly be based on a larger sample of stocks and on a longer time span than the five months we have at our disposal.

For each stock, the raw durations, denoted by X_i, are transformed into time-of-day adjusted durations (hereafter named TA-durations), denoted by x_i, by assuming the following relation:

$$x_i = X_i / \phi(t_i), \quad (1)$$

where $\phi(t_i)$ is the time-of-day effect at time t_i (i.e., the value of the estimated spline function at t_i, as detailed above). An alternative method of estimating the time-of-day function is to fit a curve by non-parametric regression of raw durations against time of day. Equation (1) corresponds to a deterministic multiplicative intra-daily seasonality function. It means that the raw duration X_i is divided by the value of the time-of-day function at time t_i, which is the "clock" time at which the duration X_i ends (since $X_i = t_i - t_{i-1}$). The resulting ratio is the TA-duration.

The time-of-day function $\phi(t_i)$ acts like a seasonal index used for seasonal adjustment of a time series. However, our time-of-day function is not normalized as is usually done in seasonal adjustment. Therefore, the scale of the adjusted durations x_i is quite different from the scale of the raw durations X_i: compare the means of the raw durations given for several stocks in Table 3 to the mean of the adjusted durations, always equal to one by construction. The fact that the adjusted durations have a mean equal to one is an advantage for the numerical maximization of the likelihood function of the models we present in Section IV (this helps to avoid overflows or underflows). The fact that we do not normalize the time-of-day function only changes the scale of the durations, but does not change their time-series properties (autocorrelations are not changed). Therefore, the maximum likelihood (ML) estimation results of the paper would not be affected (except for constant terms) if we had normalized the time-of-day function.

C. Statistical Properties of Durations

In Table 3, we report statistics on the data. For each stock and each month, we provide the number of durations (n), and the number of zero durations (n_0) that are turned into durations of 0.5 unit (given that we use two seconds as the unit). We report also the mean ($\bar{X}$), standard deviation (S), dispersion index ($S/\bar{X}$), minimum (MIN, always equal to 0.5), maximum (MAX), autocorrelation coefficient of order 1 (R_1), and Ljung-Box statistic of order 5 (Q_5) of the raw durations. The corresponding statistics (denoted with lowercase letters) for the TA-durations are given in the last three rows of each panel of the table (the mean $\bar{x}$ is not given, since it is equal to one as a result of the way the time-of-day adjustment is defined).

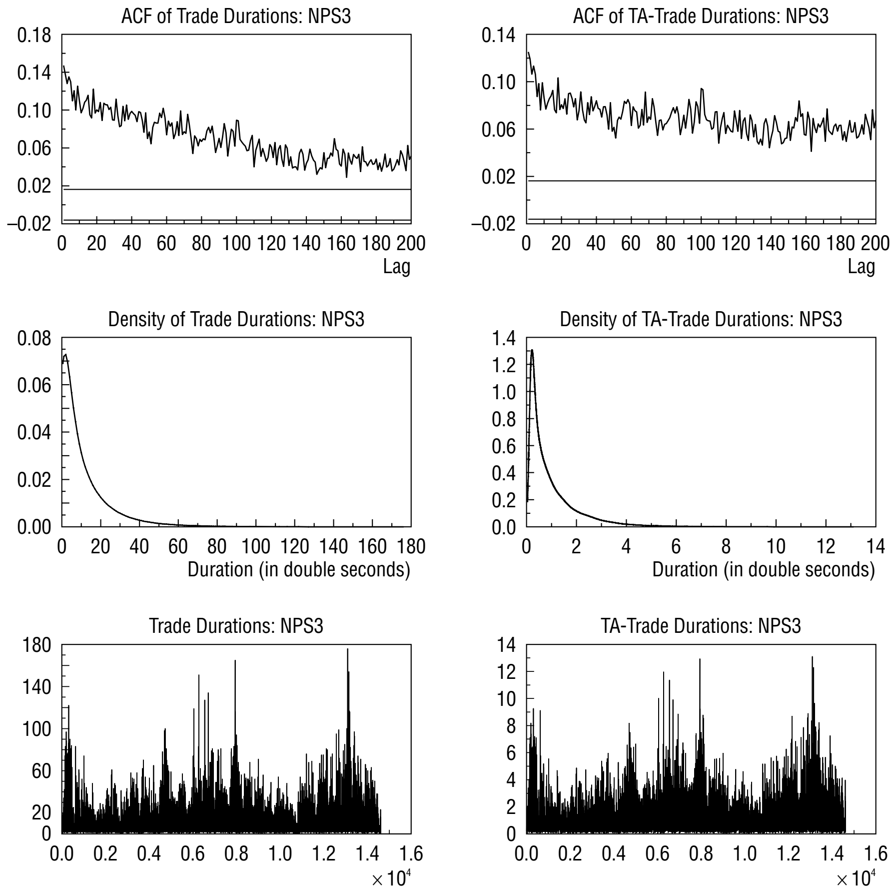
Table 3 Statistics on Trade and TA-Trade Durations

	March (20)	April (21)	May (21)	June (21)	July (22)
NPS					
n, n_0	14605, 55	13337, 30	13402, 42	18556, 109	23943, 235
$\bar{X}$, S	11.1, 12.5	12.7, 14.6	12.7, 14.2	9.2, 9.8	7.4, 7.8
$S/\bar{X}$	1.13	1.15	1.12	1.07	1.05
MIN, MAX	0.5, 176	0.5, 232	0.5, 161	0.5, 137	0.5, 117
R_1, Q_5	0.15, 1346	0.14, 887	0.09, 782	0.11, 1099	0.10, 1375
$s/\bar{x}$	1.08	1.11	1.09	1.04	1.02
min, max	0.04, 13.1	0.03, 16.4	0.03, 12.1	0.04, 16.7	0.06, 16.4
r_1, q_5	0.13, 954	0.12, 593	0.07, 543	0.09, 734	0.08, 990
SON					
n, n_0	28998, 272	36918, 440	44120, 619	39920, 938	44699, 1403
$\bar{X}$, S	5.53, 5.40	4.14, 3.89	3.83, 3.43	4.24, 3.69	3.77, 3.42
$S/\bar{X}$	0.98	0.94	0.90	0.87	0.91
MIN, MAX	0.5, 119	0.5, 70	0.5, 56	0.5, 47	0.5, 59
R_1, Q_5	0.10, 1187	0.15, 4031	0.15, 4483	0.08, 1104	0.14, 3660
$s/\bar{x}$	0.96	0.92	0.88	0.85	0.89
min, max	0.08, 28.1	0.11, 15.2	0.12, 13.6	0.10, 10.0	0.12, 14.3
r_1, q_5	0.08, 691	0.13, 3197	0.14, 3710	0.06, 490	0.11, 2648
TKE					
n, n_0	16724, 171	16871, 114	15927, 127	18893, 347	24339, 542
$\bar{X}$, S	9.66, 10.5	10.1, 11.1	10.7, 11.5	8.98, 9.72	7.31, 7.65
$S/\bar{X}$	1.09	1.10	1.08	1.08	1.05
MIN, MAX	0.5, 133	0.5, 140	0.5, 176	0.5, 112	0.5, 144
R_1, Q_5	0.08, 587	0.10, 781	0.10, 719	0.08, 833	0.09, 1242
$s/\bar{x}$	1.04	1.06	1.05	1.04	1.02
min, max	0.04, 12.3	0.04, 15.0	0.04, 14.5	0.04, 9.97	0.06, 18.6
r_1, q_5	0.06, 290	0.07, 476	0.09, 498	0.05, 354	0.07, 754
TOY					
n, n_0	29018, 305	34724, 419	29718, 287	33136, 653	33815, 903
$\bar{X}$, S	5.56, 5.51	4.89, 4.56	5.71, 5.60	5.12, 4.93	5.02, 4.89
$S/\bar{X}$	0.99	0.93	0.98	0.96	0.97
MIN, MAX	0.5, 77	0.5, 69	0.5, 62	0.5, 75	0.5, 63
R_1, Q_5	0.11, 1634	0.10, 1510	0.08, 882	0.09, 1490	0.10, 1713
$s/\bar{x}$	0.95	0.91	0.95	0.95	0.95
min, max	0.07, 12.4	0.09, 16.0	0.07, 11.4	0.08, 15.1	0.08, 10.9
r_1, q_5	0.09, 922	0.08, 1023	0.06, 471	0.07, 933	0.08, 1160

Note: n: number of durations; n_0: number of zero durations; $\bar{X}$: mean; S, s: standard deviations; MIN, min: minima; MAX, max: maxima; R_1, r_1: autocorrelation coefficients of order 1; Q_5, q_5: Ljung-Box statistics of order 5. $\bar{X}$, S, MIN, MAX, R_1, Q_5 are for raw durations, lowercase equivalent for TA-durations ($\bar{x}$ = 1 by construction). The unit for $\bar{X}$, S, MIN, MAX, min and max is two seconds. The number of trading days for each month is indicated in parentheses after the name of the month. One day of data is missing for SON and TOY in July.

To fix the interpretation, consider the NPS stock in March. There are 14,605 durations, corresponding to 730 trades per day on average, or a mean duration time of 22 seconds between two trades. The smallest duration is one second, and the longest one is almost six minutes. The first autocorrelation coefficient of the trade durations is 0.15. A graph of the autocorrelation function (ACF) is in Figure 3 (top left panel). The Q_5 statistic of 1,346 indicates strong autocorrelation after five lags, and in the figure one sees that this persists at many higher lags. For the TA-durations, the ACF starts at a lower value (top right panel), but still reveals strong dependence in the data, although the q_5 statistic is less extreme than for raw durations. This indicates that the time-of-day adjustment reduces the dependence but does not render the durations serially uncorrelated. Note that the time-of-day adjustment changes the value of the durations, but the ratio of the maximum to the minimum is not much changed (13.1/0.04 = 328, compared to 352 for the raw data).

Figure 3 ACF, Histograms, and Duration Sequences, NPS, March



Note: In the ACF graphs, horizontal lines indicate the limits of 95 percent confidence intervals ($\pm 1.96/\sqrt{n}$). Densities are estimated by a kernel method.

The statistics for the other stocks are generally similar, after accounting for the difference of general activity level (mean durations for SON and TOY are about half of those of the NPS and TKE stocks). This is even more the case if we compare the different months for the same stock. However, some differences are worth noting:

(1) The dispersion index (ratio of standard deviation to mean) is smaller than one for the two most active stocks, whereas it is larger than one for the other two stocks. Applying the test described by Engle and Russell (1998, p. 1144) for a dispersion index equal to one, this hypothesis is rejected at the level of 1 percent for all stocks and all months. This happens even if the dispersion index is equal to 1.02 as is the case for NPS and TKE in July, due to the large sample sizes (approximately 24,000 observations). These rejections reflect that the unconditional distributions of the durations are not exponential distributions.

(2) The ratio of (initially) zero to positive durations, n_0/n, is larger for the two most active stocks than for the other two.

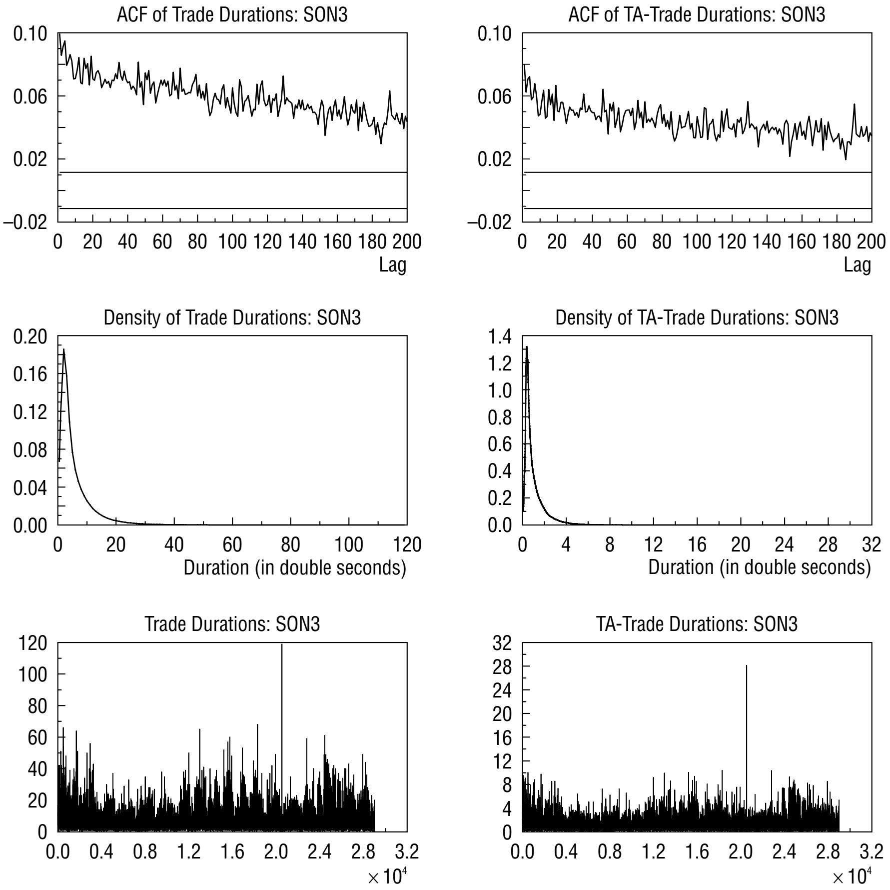
The statistical properties of the trade durations (whether TA or not) share the stylized properties also found for comparable data from other markets:

(1) Duration clustering: long durations occur in clusters, and likewise short durations. This is directly seen in the duration sequences shown in the bottom panels of Figures 3 and 4. Clustering induces positive autocorrelations and shows up in a slowly decreasing ACF that starts at a low value (between 0.05 and 0.15).

(2) Duration over/underdispersion: there are more/less extreme (small and large) durations than what is compatible with an exponential distribution. Note that previous studies have consistently reported that trade durations are overdispersed. To our knowledge, the underdispersion of the durations of SON and TOY is a specific feature of these stocks. This feature of the trade durations of SON and TOY might be an artifact due to the problem of bad measurement of the very small durations. If very small durations are underrepresented, the mean of the durations is overestimated and the standard deviation underestimated, which clearly results in an underestimated dispersion index. The densities of the durations have a narrow peak over small durations, and a long right tail (as exemplified in Figures 3 and 4, where the densities are estimated by a kernel method).

Since independent exponentially distributed durations characterize the Poisson process, the latter is not suitable to characterize trade durations. One needs a dynamic model compatible with overdispersion for trade durations.

Figure 4 ACF, Histograms, and Duration Sequences, SON, March



Note: See Figure 3 for details.

IV. Duration Analysis

In the next two subsections, we present briefly the dynamic models we use for duration analysis of the TSE data, and expose and discuss the estimation results.

A. Models

We estimated ACD models, including logarithmic versions of this model (log-ACD). There exist other models; see Bauwens *et al.* (2004) for a presentation. These authors conclude that (log-)ACD models perform at least as well as more complex ones, which are much more costly to estimate and to evaluate. Indeed, ACD models are easy to estimate by ML.

ACD models (Engle and Russell [1998]) specify the dynamics of the durations conditionally on the past durations through an autoregressive moving average structure. When modeling inter-trade durations, a first choice to make is whether to model the raw durations or the TA-durations. The first option requires including in the model the time-of-day function and maximizing jointly the likelihood function with respect to the parameters of this function and the parameters of the dynamic portion (defined below). This approach is more efficient statistically. The second option simplifies the numerical aspect of estimation without sacrificing consistency of the estimator, at the cost of some efficiency loss. In large samples like those used for the estimations reported in this paper, the efficiency loss is not likely to be a big concern, especially since the two options yield very similar results (see Engle and Russell [1998, p. 1137]). This is why we chose the second option and report estimations of ACD models for the TA-durations.

The most important assumption of ACD models is that the dependence in the duration process $\{x_i\}$ can be captured through the conditional expectation function $E[x_i|H_i]$, denoted in the sequel by Ψ_i for simplifying notation, where $H_i = \{x_{i-1}, x_{i-2}, \ldots, x_0\}$.

This is supposed to hold in such a way that $\{x_i/\Psi_i = \epsilon_i\}$ is independent and identically distributed (IID). Hence, let $\{\epsilon_i\}(i = 1, \ldots, n)$ be an IID process of positive random variables with

$$E[\epsilon_i|H_i] = E(\epsilon_i) = 1,$$

$$\mathrm{Var}(\epsilon_i|H_i) = \mathrm{Var}(\epsilon_i) = \sigma^2.$$

The ϵ_i are the "error terms" of the model.

The simplest version is the ACD (1,1) model as defined by

$$x_i = \Psi_i \epsilon_i, \tag{2}$$

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1}, \tag{3}$$

where $\omega > 0$, $\alpha > 0$, and $\beta \geq 0$ are parameters. The positivity restrictions ensure that Ψ_i cannot be negative or null, since it is the conditional mean of a positive random variable. This model may be transformed into the equivalent ARMA (1,1) model $x_i = \omega + (\alpha + \beta)x_{i-1} + u_i - \beta u_{i-1}$, where $u_i = x_i - \Psi_i$ has conditional expectation equal to zero. This implies directly that the condition $\alpha + \beta < 1$ must hold for x_i to be covariance-stationary, and that $\alpha + \beta$ is the autoregressive coefficient, which gives a measure of the persistence of the process. What the ARMA equation above also shows is that the autoregressive and moving average parameters differ only due to the parameter α. Therefore, from the properties of ARMA (1,1) models (see, e.g., Hayashi [2000, chapter 6]), we can deduce that the ACF of the process starts at a value slightly larger than α and then decays geometrically at the rate $\alpha + \beta$. With α small (i.e., between 0.05 and 0.1), and $\alpha + \beta$ large (between 0.9 and 1), this type of ACF matches the empirical ACF of inter-trade durations rather well.

The similarity of the ACD (1,1) model to the generalized autoregressive conditional heteroskedasticity (GARCH) (1,1) model of Bollerslev (1986) is obvious.

The ACD (1,1) model allows for conditional overdispersion (when $\sigma^2 > 1$) as well as underdispersion ($\sigma^2 < 1$), since the conditional dispersion index, defined as the ratio of the conditional standard deviation to the conditional mean of x_i, is equal to σ. The unconditional dispersion index, defined as the ratio $\sqrt{\mathrm{Var}(x_i)}/[\mathrm{E}(x_i)]$, is always greater than σ unless $\alpha = \beta = 0$ (for a proof, see Bauwens and Giot [2001, pp. 71–72]). The intuition of the result is that positive autocorrelation (α and $\beta > 0$) creates sequences of long durations and of short durations to a greater degree than if the durations were independent (i.e., $\alpha = \beta = 0$). These sequences increase the dispersion compared to the case of independence. Therefore, autocorrelation increases the dispersion index of ϵ_i.

By using the exponential transformation, the log-ACD class of model avoids the need for positivity restrictions on the parameters to ensure positivity of Ψ_i (this was the main motivation behind the introduction of this model by Bauwens and Giot [2000]). The version we use is the log-ACD (1,1) of type 2, taken from the reference just cited:

$$x_i = e^{\psi_i}\epsilon_i, \tag{4}$$

$$\psi_i = \omega + \alpha\epsilon_{i-1} + \beta\psi_{i-1} = \omega + \alpha\frac{x_{i-1}}{e^{\psi_{i-1}}} + \beta\psi_{i-1}. \tag{5}$$

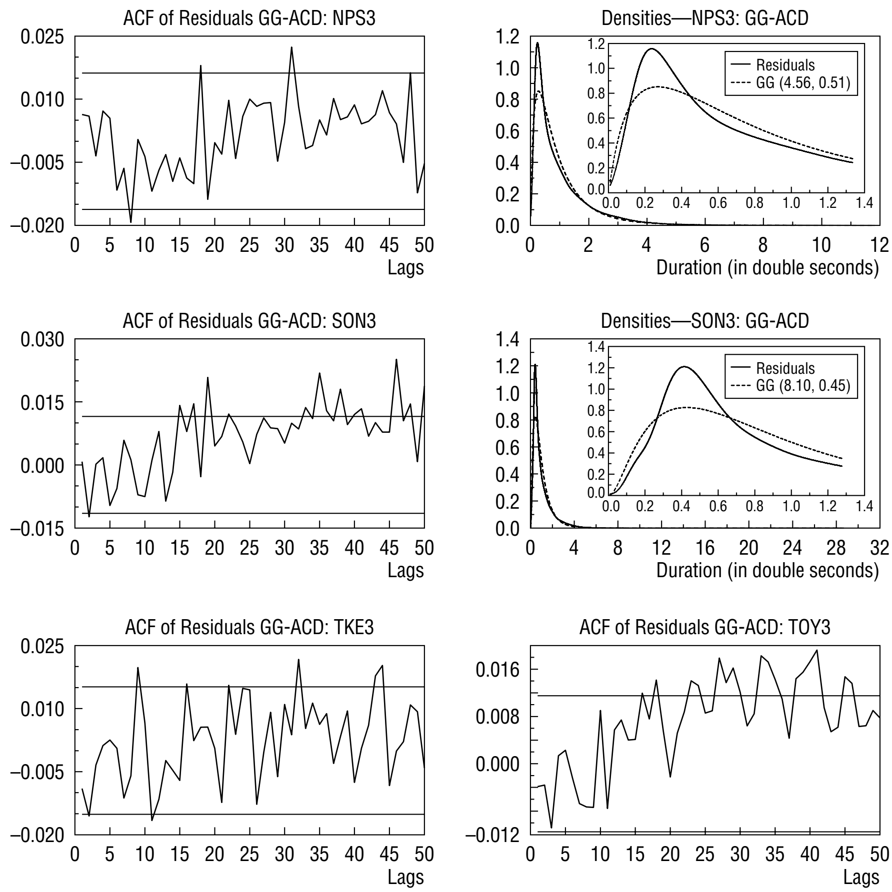
In this model, the dynamics bear on the logarithm of Ψ_i. Unconditional moments and stationarity conditions of x_i are therefore not so easy to obtain as in the ACD model. One important necessary condition for stationarity is $\beta < 1$. The parameter β corresponds to $\alpha + \beta$ (rather than to β) in the ACD model. Other conditions and expressions of moments are given in Bauwens and Giot (2001, chapter 3). The model has about the same properties as the ACD model; for example, larger persistence or autocorrelation (corresponding to larger values of β) also increases the dispersion index. However, there some minor differences: for example, Bauwens, Galli, and Giot (2003) show that the ACF decreases at a slower rate than β at low lag orders, and that this feature is sometimes more in agreement with the shape of empirical ACF functions of durations than the geometric rate of decline implied by the ACD model.

ACD and log-ACD models are flexible enough to fit the stylized properties of durations. Their specification must be completed by an assumption on the distribution of the error term ϵ_i, if ML estimation is used. We use the assumption that the distribution of ϵ_i is GG. The density is

$$f_{GG}(\epsilon_i) = \frac{\gamma}{c^{\nu\gamma}\Gamma(\nu)}\epsilon_i^{\nu\gamma-1}\exp\left[-\left(\frac{\epsilon_i}{c}\right)^{\gamma}\right], \tag{6}$$

where $\nu > 0$, $\gamma > 0$, and $c > 0$ are parameters, and $\Gamma(\nu) = \int_0^\infty u^{\nu-1}e^{-u}du$ is the gamma function. All moments exist. In particular, if $c = \Gamma(\nu)/\Gamma(\nu + \gamma^{-1})$, then $\mathrm{E}(\epsilon_i) = 1$, hence we use this expression for c and simply write $\epsilon_i \sim GG(\nu, \gamma)$. This dispersion index can be larger or smaller than one, depending on values of ν and γ (see Table 3.A.1 in Bauwens and Giot [2001, p. 100]). This is important in our context, since the duration series we handle are not always overdispersed. The GG density encompasses as particular cases the Weibull density (when $\nu = 1$), the gamma density (when $\gamma = 1$), and the exponential density (when $\nu = \gamma = 1$). Examples of GG densities are shown in Figure 5 for values of γ and ν obtained by estimation of some models. The two displayed densities are underdispersed, with dispersion indices equal to 0.96 for the GG (4.56, 0.51) and 0.79 for the GG (8.10, 0.45).

Figure 5 ACF and Densities of Residuals and Model, March



Note: Horizontal lines in the ACF graphs indicate the limits of 95 percent confidence intervals ($\pm 1.96/\sqrt{n}$). Densities of residuals are estimated by a kernel method. GG density is evaluated at the parameter estimates. Density duration ranges from zero to largest residual.

Another flexible candidate distribution for ϵ_i is the Burr. Its density is

$$f_B(\epsilon_i) = \frac{\gamma}{c}\left(\frac{\epsilon}{c}\right)^{\gamma-1}\left[1 + \lambda\left(\frac{\epsilon_i}{c}\right)^{\gamma}\right]^{-(1+\lambda^{-1})}, \tag{7}$$

where $\gamma > 0$, $\lambda > 0$, and $c > 0$ are parameters. By setting $c = \lambda^{1+\gamma^{-1}}\Gamma(1 + \lambda^{-1})/[\Gamma(1 + \gamma^{-1})\Gamma(\lambda^{-1} - \gamma^{-1})]$, the mean is equal to one, and the density has two free parameters (like the GG density). However, not all its moments exist (e.g., the mean exists if $\gamma > \lambda$). The dispersion index (when it exists) can be below or above one. The Weibull density obtains as a particular case when λ tends to zero, which is on the boundary of the admissible values, something inconvenient. Since our estimation results for the TSE data show that Burr-ACD models do not pass residual autocorrelation tests as well as GG-ACD models, we refer the reader to the appendix in chapter 3 of Bauwens and Giot (2001) for more details on the properties of the Burr distribution.

Denoting generally by $f_\epsilon(\epsilon_i;\ \theta_2)$ the density function of ϵ_i, which can depend on some parameters θ_2 (equal to (γ, ν) for the GG density, and to (γ, λ) for the Burr), the conditional density of x_i is $f_x(x_i|H_i;\ \theta) = f_\epsilon(x_i/\Psi_i;\ \theta_2)\Psi_i^{-1}$. Then the log-likelihood function (LLF) for $\theta = (\theta_1,\ \theta_2)$ where θ_1 collects the parameters of Ψ_i (typically ω, α, and β) is

$$l(\theta) = \sum_{i=1}^{n} \ln f_x(x_i|H_i;\ \theta) = \sum_{i=1}^{n}\left[\ln f_\epsilon\left(\frac{x_i}{\Psi_i};\ \theta_2\right) - \ln\Psi_i\right]. \tag{8}$$

In this expression, Ψ_i corresponds to (3) in the ACD (1,1) model, or to exp ψ_i with ψ_i as defined in (5) in the log-ACD (1,1) specification. Numerical maximization of the LLF delivers the maximum likelihood estimator (MLE) $\hat{\theta}$ with the usual large-sample approximate normality property $\hat{\theta} \sim N(\theta^0, V(\theta^0))$, where θ^0 is the so-called true value, and $V(\theta^0)$ is the variance-covariance matrix. The latter can be estimated consistently as -1 times the inverted Hessian matrix of the LLF at the maximum. Therefore, one can conduct asymptotically valid hypothesis tests on θ by usual standard normal and χ^2 tests.

B. Estimation Results

In Table 4, we provide estimation results and diagnostics for all stocks in March, and in Table 5 for the NPS stock in all months (the second columns of both tables are identical). These results are representative of what is obtained for other stock-month pairs. We report results for the GG-ACD (ACD with GG density) and ELACD2 (type 2 log-ACD with exponential density) models for reasons explained below. For each model-stock-month combination, we report ML estimates (with standard errors) and residual diagnostics (mean, dispersion index, *TED*, S_k, for k = 1, 5, 10, 50). S_k is the ratio of the Ljung-Box Q-statistic of order k of the residuals, to the 95 percent quantile of the $\chi^2(k)$ distribution. A value above one therefore indicates significant autocorrelation of order k at the 5 percent level, while a value smaller than one indicates the reverse. To complement this information, several panels in Figure 5 show the plot of the autocorrelation coefficients up to order 50 with the 95 percent confidence band. Checking residual autocorrelation is an important criterion for model evaluation: estimation is based on the assumption of IID errors, therefore significant autocorrelation in the estimated residuals is an indication of a misspecified

Table 4 Estimation Results, March, All Stocks

	NPS3	TKE3	SON3	TOY3
GG-ACD				
ω	0.021 (0.0056)	0.031 (0.012)	0.045 (0.018)	0.049 (0.016)
α	0.057 (0.0071)	0.043 (0.0065)	0.051 (0.0084)	0.063 (0.0082)
β	0.921 (0.012)	0.925 (0.019)	0.902 (0.027)	0.886 (0.024)
γ	0.51 (0.019)	0.49 (0.023)	0.45 (0.023)	0.46 (0.021)
ν	4.56 (0.33)	4.74 (0.43)	8.10 (0.80)	7.97 (0.71)
$\bar{\epsilon}$, d	1.01, 1.00	1.01, 1.02	1.01, 0.92	1.01, 0.91
S_1, S_5	0.15, 0.22	0.35, 0.52	0.00, 0.65	0.11, 0.40
S_{10}, S_{50}	0.58, 0.84	0.86, 1.23	0.67, 2.46	0.62, 2.38
ELACD2				
ω	–0.032 (0.0046)	–0.019 (0.0026)	–0.019 (0.0021)	–0.026 (0.0032)
α	0.032 (0.0045)	0.019 (0.0026)	0.019 (0.0021)	0.026 (0.0032)
β	0.994 (0.0019)	0.995 (0.0014)	0.996 (0.0010)	0.994 (0.0015)
$\bar{\epsilon}$, d	1.00, 0.99	1.00, 1.01	1.00, 0.90	1.00, 0.90
TED	–0.54	–0.53	–11.0	–11.9
S_1, S_5	1.32, 1.71	0.07, 0.44	3.40, 2.19	2.88, 2.62
S_{10}, S_{50}	1.22, 0.98	1.03, 1.06	1.70, 1.05	1.74, 1.03
n	14,605	16,724	28,998	29,018

Note: Standard errors are reported in parentheses below the estimates. $\bar{\epsilon}$, d: mean and dispersion index (standard deviation/mean) of residuals; *TED*: N (0,1)-test statistic for overdispersion of residuals under exponential null hypothesis; S_k, $k = 1 \ldots 50$: ratio of Ljung-Box Q-statistic of order k of residuals to 95 percent quantile of $\chi^2(k)$; n: number of observations used for estimation.

Table 5 Estimation Results, NPS, All Months

	March	April	May	June	July	All
GG-ACD						
ω	0.021 (0.0056)	0.024 (0.0060)	0.026 (0.0059)	0.028 (0.0050)	0.011 (0.0021)	0.017 (0.0017)
α	0.057 (0.0071)	0.063 (0.0074)	0.051 (0.0051)	0.058 (0.0047)	0.038 (0.0033)	0.045 (0.0023)
β	0.921 (0.012)	0.913 (0.013)	0.922 (0.010)	0.913 (0.0091)	0.951 (0.0053)	0.938 (0.0039)
γ	0.51 (0.019)	0.49 (0.020)	0.51 (0.017)	0.47 (0.015)	0.52 (0.014)	0.41 (0.0088)
ν	4.56 (0.33)	4.49 (0.34)	4.12 (0.27)	5.48 (0.34)	5.13 (0.26)	7.07 (0.30)
$\bar{\epsilon}, d$	1.01, 1.00	1.01, 1.04	1.01, 1.03	1.01, 0.97	1.01, 0.94	1.01, 0.99
S_1, S_5	0.15, 0.22	0.73, 0.95	3.51, 1.33	0.25, 0.23	0.84, 0.32	0.04, 0.27
S_{10}, S_{50}	0.58, 0.84	0.76, 1.30	1.00, 0.93	0.23, 0.70	0.30, 0.64	0.33, 0.98
ELACD2						
ω	–0.032 (0.0046)	–0.035 (0.0044)	–0.033 (0.0038)	–0.041 (0.0045)	–0.021 (0.0020)	–0.030 (0.0018)
α	0.032 (0.0045)	0.035 (0.0043)	0.033 (0.0037)	0.041 (0.0044)	0.020 (0.0020)	0.029 (0.0018)
β	0.994 (0.0019)	0.994 (0.0018)	0.991 (0.0023)	0.988 (0.0026)	0.997 (0.00063)	0.995 (0.00073)
$\bar{\epsilon}, d$	1.00, 0.99	1.00, 1.02	1.00, 1.02	1.00, 0.97	1.00, 0.94	1.00, 0.98
TED	–0.54	1.76	1.64	–2.88	–6.94	–3.76
S_1, S_5	1.32, 1.71	3.48, 1.71	1.58, 1.03	0.03, 0.43	0.03, 0.59	1.37, 3.76
S_{10}, S_{50}	1.22, 0.98	1.07, 1.29	0.93, 0.77	0.36, 0.83	0.57, 0.73	2.57, 2.09
n	14,605	13,337	13,402	18,556	23,943	83,843

Note: For definitions and details, refer to Table 4.

model. *TED* is the test statistic of Engle and Russell (1998, p. 1144), asymptotically N (0,1), for the null that the dispersion index of the errors is equal to one and applies only to the ELACD2 residuals.

The set of models that were fitted to each stock-month data set include the ACD model defined by equations (2)–(3) and the log-ACD model defined by equations (4)–(5), each combined with the following distributions for ϵ_i: GG, defined by equation (6), Burr, Weibull, and exponential, making a total of eight models.

Equations (3) and (5) correspond to (1,1)-models and can be extended to include more lags of the duration and its conditional expectation (a model is said to be of order (p, q) if it includes p lags of the duration and q lags of the conditional expectation). Concerning the choice of p and q, it is almost a stylized fact that (1,1)-models provide a correct specification of the dynamics of the durations and are not dominated by models with p or q greater than one (this is similar to what happens with GARCH models). Therefore, we started our specification search (for each possible type of model) with the (1,1)-model and did not look for models of higher order if the (1,1)-model passed the residual autocorrelation checks. This does not exclude that, for example, a (2,1)-model can be correctly specified. However, in that case we prefer the most parsimonious model provided that it is not rejected by a likelihood ratio test against the more general model. In this context, the Akaike information criterion (AIC) or Bayesian information criterion (BIC) could be used as a model choice criterion (rather than the likelihood ratio test), and the final outcome may depend on the criterion used. Our specification searches resulted in the choice of (1,1)-models in all cases. A synthesis of the results follows.

(1) Across all months/stocks, the GG-ACD model is generally the overall best-fitting model, that is, it has usually better specification diagnostics than the other model/distribution combinations. This is not surprising if one compares it with models using the Weibull (or exponential) density, since these densities have one (or two) fewer parameters and are therefore much less able to fit the data. The Burr distribution has as many parameters as the GG, but typically in our results the residuals of Burr ACD models are significantly autocorrelated (at the 5 percent level) even at low lags, whereas this is not the case with GG-ACD models (with the exception of the results for NPS in May, see Tables 4 and 5). Moreover, since numerical convergence problems occur in estimating the type 2 log-ACD with GG density (GG-LACD2) models (for unknown reasons) it is not possible to compare GG-ACD with GG-LACD2 models.

(2) Many models still have autocorrelated residuals at some lags, even if autocorrelation is considerably reduced compared to what is found in the TA-durations. In this respect, GG-ACD models for TOY and SON (the most active stocks) still have strongly significant residual autocorrelation at lag 50 (see the value of S_{50} and the ACF panels of Figure 5), while this is much less true for the other stocks (NPS and TKE). This feature may indicate the need to use a model compatible with "long memory" for such series.

(3) ELACD2 model estimates are reported for comparison with the GG-ACD estimates. In terms of diagnostics, the GG-ACD models generally do a better job of removing autocorrelation from the residuals, although there are few exceptions (e.g., S_{10} and S_{50} values for NPS in May are much lower under ELACD2 than GG-ACD).

(4) Estimates of the parameters of the autoregressive equation are in the stationary region but close to the boundary, reflecting the high degree of autocorrelation of the durations. However, the sum of the estimates of $\alpha + \beta$ is more distant from one in GG-ACD models than in other ACD models (see also the β estimates very close to one in the ELACD2 results). When we use an exponential distribution, $\alpha + \beta$ is pushed toward the value one since this allows a better capture of the dispersion of the data. When parameters like those of the GG density are present, they help to capture this feature and free to some extent the dynamic parameters of this aspect. It is likely that more flexible ACD models will help to better separate the need to fit both the persistence and the dispersion of the process. The Markow-switching ACD model (Hujer, Vuletic, and Kokot [2002]) and the mixture duration model (DeLuca and Gallo [2004] and Hujer and Vuletic [2005]) have been proposed recently as valuable extensions of ACD models.

(5) In all cases, a formal statistical test of goodness-of-fit of the estimated distribution of ϵ_i rejects the assumed form of distribution at any conventional

significance level (this is the test for uniform distribution of the probability integral transforms of the durations described in Bauwens *et al.* [2004]; see formula 4). The value of the test statistic is less extreme for the GG density than for the competitors (Burr, Weibull, exponential), indicating that the GG density is the "least bad" choice. The panels of Figure 5 showing the fitted GG density and a kernel density of the residuals for NPS and SON illustrate the inadequacy of the GG density to fit the density of the residuals. Even though the shapes of the two curves are rather similar, the GG density puts too little mass in the area around the mode, and consequently too much mass on surrounding areas. The discrepancy is larger for SON than for NPS. Bauwens *et al.* (2004) report similar results for NYSE trade durations.

(6) For GG-ACD models, the dispersion index of the residuals is close to one in most cases. It is smaller than one for SON and TOY in March, where it takes a value close to 0.91, and for NPS in July (0.94) and June (0.97). The results are very similar for ELACD2 models. The *TED* statistic in this case reveals that the dispersion index is significantly different from one for SON and TOY in March, and for NPS in June and July. Although the *TED* statistic is not applicable to the GG density, one may presume that the dispersion index values for SON and TOY in March and NPS in July and June mentioned above for GG-ACD models are also significantly different from one. Notice that finding (as in the other cases) that they are not significantly different from one does not imply that the exponential distribution is the best choice.

(7) Estimates for different months on the same stock are rather stable. (See Table 5 for the NPS stock.) This indicates that the process generating the trades was stable over the period of March–July 2003. For comparison, we also report the estimates for the pooled data sets of NPS (five months altogether); see the last column of Table 5. Note the acceptable dynamic specification of the GG-ACD model in this case (no S_k value is larger than one). A likelihood ratio test of equality of the parameters for the five months rejects the null hypothesis at any conventional level, but this is hardly surprising given the large sample size. The most different parameters are those of the GG density.

(8) We also estimated the models when the raw durations that are equal to zero seconds are removed rather than replaced by durations equal to one. The previous comments remain valid, since the estimation results are not much changed. The most sensitive estimates are the parameters of the GG density. This is not surprising, since these parameters serve to fit the distribution and the removal of observations occurs mainly at the left tail of the distribution. For example, the estimates for the TKE stock in March are $\hat{\nu} = 5.29$ and $\hat{\gamma} = 0.47$, instead of 4.74 and 0.49. For SON in March, they are 8.86 and 0.44, instead of 8.10 and 0.45. Such changes are of course less important when the number of zero durations (turned into one) is very small (as for NPS in March). The most interesting difference is that residual autocorrelation is generally more important for GG-ACD models when the zero durations are removed from the data.

V. Conclusion

The main purpose of this paper is to characterize the statistical properties of inter-trade durations of four stocks traded on the TSE, and to estimate econometric dynamic duration models that fit the data. We find that the GG-ACD (1,1) model captures correctly the dynamic properties of the durations, although for two stocks there remains some significant autocorrelation (at order 20 or higher) in the residuals. We stress that despite this shortcoming the degree of autocorrelation of the residuals is very much lower than what it is for the durations (a similar result is reported by Engle and Russell [1998] for the trade durations of the stock of International Business Machines Corp. [IBM] on the NYSE), indicating the usefulness of the ACD model.

It should be borne in mind that the data we analyzed pertain to highly traded stocks and that the properties found for these data do not extend necessarily to much less frequently traded stocks. Not surprisingly, the TSE data share the main properties of comparable data for other markets such as the NYSE, but we found some small differences, which are linked to the specificity of the TSE and to the imperfect nature of the recording of the data. We think that more in-depth studies of the TSE market would benefit from more precise and extensive data recording. Table 6 provides, in qualitative terms, a comparison of the main features of trade durations and estimated ACD-type models for TSE and NYSE stocks.

Table 6 Summary of TSE and NYSE Trade Durations and ACD Models

	TSE	NYSE
Durations		
—Intra-daily pattern	Inverted-U	Same
—Dispersion index	Between 0.85 and 1.10	Larger than one
—Autocorrelations	Start around 0.10	Same
	Decrease slowly	Same
—Density mode	About 0.3	Same
—Density shape	Very sharp around mode	Same
	With long right tail	Same
Best-fitting models		
—Type	ACD	ACD and log-ACD
—Persistence	High	High
—Distribution	GG	Same or Burr

Note: The qualitative features pertain to highly traded stocks. For details about trade durations of NYSE stocks, see Bauwens and Giot (2001). Duration features, except the intra-daily pattern, are for time-of-day adjusted durations.

References

Bauwens, L., F. Galli, and P. Giot, "The Moments of Log-ACD Models," CORE Discussion Paper No. 2003/11, Université catholique de Louvain, 2003.

———, and P. Giot, "The Logarithmic ACD Model: An Application to the Bid/Ask Quote Process of Two NYSE Stocks," *Annales d'Economie et de Statistique*, 60, 2000, pp. 117–149.

———, and ———, *Econometric Modelling of Stock Market Intraday Activity*, Dordrecht: Kluwer, 2001.

———, ———, J. Grammig, and D. Veredas, "A Comparison of Financial Duration Models via Density Forecasts," *International Journal of Forecasting*, 20, 2004, pp. 589–609.

———, and D. Veredas, "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations," *Journal of Econometrics*, 119, 2003, pp. 381–412.

Bollerslev, T., "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 1986, pp. 307–327.

DeLuca, G., and G. Gallo, "Mixture Processes for Financial Intradaily Durations," *Studies in Nonlinear Dynamics and Econometrics*, 8, 2004, pp. 1–18.

Engle, R. F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 1982, pp. 987–1007.

———, and J. R. Russell, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1998, pp. 1127–1162.

Ghysels, E., C. Gouriéroux, and J. Jasiak, "Stochastic Volatility Duration Models," *Journal of Econometrics*, 119, 2004, pp. 413–433.

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

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

Hayashi, F., *Econometrics*. Princeton: Princeton University Press, 2000.

Hujer, R., and S. Vuletic, "Econometric Analysis of Financial Trade Processes by Mixture Duration Models," working paper, University of Frankfurt/Johann Wolfgang Goethe University, 2005.

———, ———, and S. Kokot, "The Markov-Switching ACD Model," Finance and Accounting Working Paper No. 90, Johann Wolfgang Goethe University, 2002.

Lancaster, T., *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press, 1990.

Lunde, A., "A Generalized Gamma Autoregressive Conditional Duration Model," discussion paper, Aalborg University, 2000.

Madhavan, A., "Market Microstructure: A Survey," *Journal of Financial Markets*, 3, 2000, pp. 205–258.

O'Hara, M., *Market Microstructure Theory*, Oxford: Basil Blackwell, 1995.

Tokyo Stock Exchange, "Guide to TSE Trading Methodology," 2nd edition, 2003 (available at http://www.tse.or.jp/english/cash/stock/methodology/guide.pdf).

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