MODELLING INTRADAY TRADING ACTIVITY USING BOX-COX ACD MODELS

NIKOLAUS HAUTSCH[*]

CENTER OF FINANCE AND ECONOMETRICS,
UNIVERSITY OF KONSTANZ

NIKOLAUS.HAUTSCH@UNI-KONSTANZ.DE

ABSTRACT. In this paper, I model the intraday trading activity based on volume durations, i.e. the waiting time until a predetermined volume is absorbed by the market. Since this concept measures the trading volume per time it is strongly related to market liquidity. I focus on volumes measured independently of the side of the market as well as on buy volumes, sell volumes and volumes measured on both market sides simultaneously. For econometric modelling of the different duration concepts, the performance of alternative types of Box-Cox-ACD models are analyzed. By evaluating out-of-sample forecasts, evidence is provided that Box-Cox-ACD models are a valuable tool for predicting volume durations. It is shown that volume durations measured independently of the side of the market have the best predictability. Furthermore, I illustrate that the inclusion of explanatory variables capturing past market activities concerning the price process and imbalances between the buy and sell side of the market. The empirical study uses IBM transaction data from the NYSE.

1. INTRODUCTION

This paper analyzes the time and volume dimension of the intraday trading process. The main idea is to investigate volume durations, i.e. the waiting time until a predetermined volume is traded on the market. Since volume durations measure the speed of the market with respect to the trading volume, i.e. the trading volume per time, they are a valuable proxy for liquidity. The main advantage of this liquidity concept is that it is easily derived from the trade and quote process of a market and does not require insights into the limit order book.

An economically quite reasonable interpretation of volume durations is provided by Gourieroux, Jasiak, and LeFol (1999). They interpret the waiting time until a given volume is traded as liquidity costs, i.e. as the (time) costs, a trader is faced with when his order is not immediately executed. Moreover, by accounting for the type of the corresponding trades it is possible to focus on different aspects of the intraday trading process. E.g. the analysis of buy (sell) volumes provides deeper insights into the trading activities on the particular sides of the market and yield indications concerning the capacities of the order book. In this sense volume durations might be associated with execution times of unlimited market orders. Furthermore, the analysis of the waiting time until a given volume is traded on both sides of the market allows not only to account for the speed of the market but also for the balance between the market sides.

Therefore, in this study, three different types of volume durations are analyzed: I investigate the waiting time until a predetermined aggregated volume is traded, (i) independent from the market side, (ii) on the buy (sell) side, and (iii) on both market sides simultaneously. From an economic point of view two main questions are resolved within the course of the paper. First, how predictable are volume durations and thus the absorptive capacities of the market, especially on the particular market sides? Second, which impact have past market activities on the expected volume duration and thus the speed of the market? In particular, are volatile market periods followed by more or less liquid phases? Which impact have large price movements and are there asymmetry effects with respect to the trader's behaviour on the different sides of the market?

The econometric framework to estimate volume durations is provided by the seminal work of Engle and Russell (1998), who proposed the Autoregressive Conditional Duration (ACD) model which shows a strong resemblance to the GARCH model for price processes. In this paper, the performance of alternative ACD specifications with respect to the goodness-of-fit and the prediction of volume durations is analyzed. From an econometric point of view I focus on two major aspects: First, which functional form of the conditional mean function of the ACD model is suitable to model different types of volume durations and provides satisfying forecasts? Second, which distributional assumptions are appropriate?

Therefore, in this paper the ACD framework is extended in two directions. First, more flexible functional forms of the conditional mean function based on Box-Cox transformations are proposed. These new types of Box-Cox-ACD models are quite flexible and nest the basic ACD model (Engle and Russell, 1998), the Log-ACD model (Bauwens and Giot, 2000) and the Box-Cox-ACD model proposed by Dufour and Engle (2000). Second, as

proposed by Hautsch (2001), the ACD error term follows a Generalized F distribution which allows for a wide variety of different shapes of the hazard function.

Based on out-of-sample predictions of the mean and of the density, it will be shown that more flexible specifications of the conditional mean function improve both the fit and the forecast power of the models. Contrarily, more flexible distributions improve the fit of the data, but do not significantly affect predictions of the duration mean and the duration density. Furthermore, it is shown that explanatory variables associated with trading activities within the last 10 minutes have a significant impact on the expected volume duration, even when dynamic dependencies are taken into account. Moreover, they improve the predictability of the mean.

The paper is organized in the following way: In section 2, a characterization of volume durations and their relationship to liquidity measures is discussed in more detail. Section 3 presents different types of (Box-Cox-) ACD models based on the Generalized F distribution. Section 4 deals with data description, the derivation of the different volume duration concepts and the presentation of summary statistics. Section 5 gives the estimation results while section 6 discusses evaluations of the prediction performance of the different ACD specifications. In section 7, the impact of explanatory variables associated with past market activities is analyzed. The conclusions are given in section 8.

2. Volume durations and liquidity

Liquidity has been recognized as an important determinant of market behaviour and the efficient working of a market. Following the conventional definition of liquidity (see e.g. Keynes, 1930, Demsetz, 1968, Black, 1971, Glosten and Harris, 1988) an asset is considered as liquid if it can be traded quickly, in large quantities and with little impact on the price. Thus, following this concept, the measurement of liquidity requires to account for three dimensions of the transaction process: Time, volume and price. Kyle (1985) defines liquidity in terms of the tightness indicated by the bid-ask spread, the depth corresponding to the amount of one sided volume that can be absorbed by the market without inducing a revision of the bid and ask quotes and resiliency, i.e. the time in which the market returns to its equilibrium. The multidimensionality of the liquidity concept is also reflected in theoretical and empirical literature, where several strings can be divided: A wide range of the literature is related to the bid-ask spread as a measure of liquidity[1] and to the decomposition of the spread with a main focus on the measurement of the adverse

---

[1]See e.g. Bessembinder (2000), Elyasiani, Hauser, and Lauterbach (2000), Greene and Smart (1999) or Conroy, Harris, and Benet (1990).

selection cost component[2]. Other studies deal with the analysis of the market depth and the order flow in the limit order book[3].

A further string of literature is related to the analysis of trading volume[4] and its price impact[5]. In electronic trading systems, the price impact is determined by the market depth, i.e. the absorptive capacities of the order queues in the limit order book. Thus, the larger the volume an investor wants to buy or sell, the higher the probability that it exceeds the capacity of the first queue of the limit order book and thus the larger is the price impact. If the demand (or supply) is not large enough to match the order, the investor has to wait until execution is guaranteed. In a market maker market the price impact is determined by the order book of the market maker and thus the posted bid-ask spread. The larger the volume a trader wants to buy or to sell, the larger the spread posted by the market maker in order to account for his adverse selection risk and his inventory costs. Thus, the investor has to bear liquidity costs which arise through the difference between the market price and the ask (bid) quote placed by the market maker. Since the order book of the market maker is unobservable, the price impact of a large volume in each instant of time is hardly identifiable.

Quite natural liquidity measures arise by ignoring the price impact and focussing on the time dimension of the intraday trading process. In this context, intertrade durations are a proxy for the arrival rate of new orders[6] and essential determinants of market liquidity. A valuable way not only to account for the time dimension but also for the volume dimension is to consider volume durations, i.e. the time in which a certain volume is absorbed by the market[7]. Even though volume durations do not account for the price impact, they admit a quite reasonable interpretation as time costs of liquidity (see Gourieroux, Jasiak, and LeFol, 1999). Consider a trader who wants to execute a large order but wants to avoid the costs for immediacy induced by a high bid-ask-spread. Then he has the possibility to split his order and to distribute the volume over time[8]. Such a trader is interested in the time he has to wait until the execution of the complete order. Then, the expected volume duration allow him to quantify the (time) costs of liquidity.

---

[2]See Huang and Stoll (1997), George, Kaul, and Nimalendran (1991) or Glosten (1987) among others.
[3]See e.g. Glosten (1994), Biais, Hillion, and Spatt (1995) or Pirrong (1996).
[4]See Aggarwal and Gruca (1993), Kim and Verrecchia (1994), Foster and Viswanathan (1993) or Easley, Kiefer, O'Hara, and Paperman (1996).
[5]See e.g. Chan and Lakonishok (1995), Keim and Madhavan (1996) or Fleming and Remolona (1999).
[6]See Easley and OHara (1991), Kluger and Stephan (1997), Hautsch (1999) or Al-Suhaibani and Kryzanowski (2000).
[7]See e.g. Engle and Lange (1997) or Gourieroux, Jasiak, and LeFol (1999).
[8]Such a behaviour might also be reasonable due to strategic reasons.

By defining volume durations not only based on the amount of volume shares but also on the type of the corresponding transactions, it is possible to capture different components of the trading process. Hence, buy (sell) volume durations might be interpreted as the waiting time until a corresponding unlimited market order is executed. In this sense, forecasts of volume durations are associated with predictions of the absorptive capacities of the market, especially for the particular market sides. Alternatively, by the measurement of the time until a given volume on both market sides is traded, one obtains a liquidity measure which also accounts for the balance between the market sides. Then, a market period is defined as liquid if unlimited market orders are executed quickly on both sides of the market.

Hence, volume durations allow to focus on different aspects of liquidity and might be used as valuable means to compare different market scenarios as well as different markets.

3. ACD Models

Let τ_t, $t = 1, \ldots, T$, denote the waiting time until a given volume v is absorbed by the market. The conditional expected volume per time is given by

$$\rho = \frac{\mathrm{E}\left[\tau_t | \mathcal{I}_{t-1}\right]}{v}, \tag{1}$$

where $\mathcal{I}_t$ denotes the information set up to period t. The function $\mathrm{E}\left[\tau_t | \mathcal{I}_{t-1}\right]$ is easily estimated by applying the ACD framework proposed by Engle (1996)[9] and Engle and Russell (1998) to model point processes with dependent arrival rates. The main principle of the ACD model is to specify the durations τ_t as a multiplicative relationship between the conditional mean function $\Psi_t = \mathrm{E}[\tau_t | \mathcal{I}_{t-1}]$ and an error term ϵ_t with positive support. Thus

$$\tau_t = \Psi_t \cdot \epsilon_t \quad \text{with } \mathrm{E}[\epsilon_t] = 1. \tag{2}$$

Different types of ACD models can be divided either by choice of the functional form for the conditional mean function Ψ_t or by choice of the distribution for ϵ_t.

The basic ACD specification proposed by Engle and Russell (1998) and Engle (2000) is based on a linear parameterization of the conditional mean function[10]

$$\mathrm{ACD(p,q):} \ \Psi_t = \omega + \alpha_1 \tau_{t-1} + \beta_1 \Psi_{t-1}. \tag{3}$$

Dufour and Engle (2000) discuss two main drawbacks of this specification: First, this formulation requires constraints on the parameters to ensure that the model does not predict negative durations. Second, they provide evidence that non-linear functional forms

---

[9] The paper is now published as Engle (2000).

[10] Note that for simplicity the lag order is set to one.

of Ψ_t are more appropriate to model the adjustment process of the conditional mean to recent durations.

A valuable alternative which requires no parameter constraints is to specify the conditional mean function Ψ_t in logarithmic form, leading to the Log-ACD model proposed by Bauwens and Giot (2000)

$$(4) \quad \text{Log-ACD(p,q):} \ \ln \Psi_t = \omega + \alpha_1 \ln \tau_{t-1} + \beta_1 \ln \Psi_{t-1}$$

$$= \omega + \alpha_1 \ln \epsilon_{t-1} + \tilde{\beta}_1 \ln \Psi_{t-1} \qquad \text{with } \tilde{\beta}_1 = \alpha_1 + \beta_1.$$

However, this specification implies a relatively rigid adjustment process of the conditional mean to recent durations and thus, in general, an overadjustment of the conditional mean after very short durations. Therefore, Dufour and Engle (2000) propose the Box-Cox-ACD model as a more flexible alternative based on a Box-Cox transformation of the past innovations,

$$(5) \quad \text{BC-ACD(p,q):} \ \ln \Psi_t = \omega + \alpha_1 (\epsilon_{t-1}^\delta - 1)/\delta + \beta_1 \ln \Psi_{t-1}.$$

Engle and Dufour's Box-Cox-ACD specification includes the Log-ACD model for the Box-Cox parameter $\delta \to 0$ and a linear specification for $\delta = 1$.

The major argument for the use of a logarithmic form for Ψ_t is to ensure the non-negativity condition of the predictions of the model. However, note that the functional form of Ψ_t has important implications for the marginal impact of past durations on the current duration since a logarithmic form implies a multiplicative relationship between past durations which is quite different from a linear form. The crucial question is whether the restriction on the functional form of Ψ_t because of non-negativity conditions is justified. In order to allow for a higher flexibility, I propose a specification which weakens the non-negativity condition, but allows to test a linear form of the conditional mean function against a logarithmic one. The main idea is to specify the random variable Ψ_t itself in terms of a Box-Cox transformation leading to

$$(6) \quad \text{BC1-ACD(p,q):} \ (\Psi_t^\delta - 1)/\delta = \omega + \alpha_1 (\epsilon_{t-1}^\delta - 1)/\delta + \beta_1 (\Psi_{t-1}^\delta - 1)/\delta.$$

This specification includes the linear ACD model (in terms of past innovations) for $\delta = 1$ and the Log-ACD model for $\delta \to 0$. The model is rewritten as

$$\text{BC1-ACD(p,q):} \ \Psi_t^\delta = \tilde{\omega} + \alpha_1 \epsilon_{t-1}^\delta + \beta_1 \Psi_{t-1}^\delta,$$

where $\tilde{\omega} = \omega + 1 - \alpha_1 - \beta_1$. Hence, this ACD specification is the counterpart to the

Higher flexibility is achieved by the inclusion of two Box-Cox parameters

$$(7) \quad \text{BC2-ACD(p,q):} \ (\Psi_t^{\delta_1} - 1)/\delta_1 = \omega + \sum_{j=1}^{p} \alpha_j (\epsilon_{t-j}^{\delta_2} - 1)/\delta_2 + \sum_{j=1}^{q} \beta_j (\Psi_{t-j}^{\delta_1} - 1)/\delta_1.$$

The advantage of this specification is that it nests the BC1-ACD model, Engle and Dufour's BC-ACD model as well as the Log-ACD model and the basic ACD model. Despite of the high nonlinearity of this model induced by the different Box-Cox transformations, it is easily estimated by ML without imposing any parameter restrictions.

A further possible extension of this framework would be to allow for asymmetric news impact curves as proposed by Fernandes and Grammig (2001). In this framework the model is built based on an absolute value function of the past innovations leading to non-continuous news impact curves. However, even though such a specification allows for quite flexible news impact curves, the authors remark that it is not easily estimated since the inclusion of absolute value functions lead to numerical problems when the log likelihood function is maximized. Moreover, in most of the cases the estimation of the Hessian is quite cumbersome and OPG standard errors are computed instead of the robust sandwich form.

Therefore, in order to avoid such problems and make the inference of the particular specifications comparable, this study is restricted to the models given above which allow, even based on the Generalized F distribution, for a trouble-free estimation and for robust inferences.

Stationarity is ensured for the basic ACD model by $\sum_{j=1}^{p} \alpha_j + \sum_{j=1}^{q} \beta_j \leq 1$ and for the Log-ACD model as well as the Box-Cox ACD specifications by $\sum_{j=1}^{q} \beta_j \leq 1$ (see also Engle and Russell, 1998 or Dufour and Engle, 2000).

Focussing on the choice of the distribution of the error term ϵ_t, the most obvious choice is the standard exponential distribution which is a relatively restrictive parameterization for most of the applications. More flexible distributions are the Weibull distribution (Engle and Russell, 1998), the Generalized Gamma distribution (Lunde, 2000), the Burr distribution (Grammig and Maurer, 2000) or the Generalized F distribution (Hautsch, 2001). In order to get deeper insights into the role of distributional flexibility, I use the Generalized F distribution and as benchmark the Weibull distribution.

The mean of the Generalized F distribution is given by

$$(8) \quad \text{E}[\tau] = \lambda^{-1} \eta^{1/a} \frac{\Gamma(m + 1/a)\Gamma(\eta - 1/a)}{\Gamma(m)\Gamma(\eta)}, \quad a\eta > 1,$$

where a, m and η are parameters determining the shape of the hazard function and λ is a scale parameter. As outlined by Hautsch (2001), the Generalized F ACD model is built

on a time dependent specification of the inverse of the scale parameter $\theta := \lambda^{-1}$, i.e. it is assumed that

$$(9) \qquad\qquad \Psi_t = \theta_t \zeta,$$

where

$$(10) \qquad\qquad \zeta := \frac{\eta^{1/a}\Gamma(m+1/a)\Gamma(\eta-1/a)}{\Gamma(m)\Gamma(\eta)}.$$

This model includes as special cases the generalized gamma distribution for $\eta \to \infty$, the Weibull family for $m = 1, \eta \to \infty$ and the log-logistic distribution for $m = \eta = 1$. Combining this specification with the BC2-ACD model, eq. (7), gives

$$(11) \qquad ((\theta_t\zeta)^{\delta_1} - 1)/\delta_1 = \omega + \sum_{j=1}^{p} \alpha_j(\tau_{t-j}^{\delta_2} - 1)/\delta_2 + \sum_{j=1}^{q} \beta_j((\theta_{t-j}\zeta)^{\delta_1} - 1)/\delta_1.$$

The log likelihood function is obtained by

$$(12) \qquad \log\mathcal{L} = \sum_{t=1}^{T} \log\frac{\Gamma(m+\eta)}{\Gamma(m)\Gamma(\eta)} + \log a - am\log\theta_t$$
$$+ (am-1)\log\tau_t - (\eta+m)\log\left[\eta + (\tau_t/\theta_t)^a\right].$$

A specific feature of financial durations is a strong impact of intraday seasonality patterns, see e.g. Wood, McInish, and Ord (1985), Engle and Russell (1998), Giot (2000) or Gerhard and Hautsch (2001). One common solution within the ACD framework is to generate seasonally adjusted series by partialling out the time-of-day effects. In this context, the durations are decomposed into a deterministic and a stochastic component. Engle and Russell (1998) formulate the deterministic seasonality effect as a multiplicative function, which is given by

$$(13) \qquad\qquad \tau_t = \tilde{\tau}_t s_t(\tau_{t-1}, \xi),$$

where $s_t(\tau_{t-1}, \xi)$ corresponds to the seasonality function depending on seasonality parameters ξ associated with the beginning of the spell, and $\tilde{\tau}_t$ denotes the 'seasonal adjusted' duration. Thus, in this context the conditional expectation of τ_t is

$$(14) \qquad\qquad \mathrm{E}\left[\tau_t | \mathcal{I}_{t-1}\right] = \tilde{\Psi}_t s_t,$$

where $\tilde{\Psi}_t$ denotes the conditional mean function of the seasonal adjusted durations.[11]

---

[11]An alternative seasonal adjustment procedure is proposed by Veredas, Rodriguez-Poo, and Espasa (2001) who specified a semiparametric estimator where the seasonal components are jointly estimated non-parametrically with the parameters of the ACD model.

4. CONSTRUCTING LIQUIDITY MEASURES BASED ON VOLUME DURATIONS

I use the IBM data used in Engle and Russell (1998) and Engle (2000) which is extracted from the Trades and Quotes (TAQ) database available from the NYSE. Trading at the NYSE is based on a so-called hybrid system, i.e. the trading mechanism combines a market maker system with an order book system. For each stock one market maker (specialist) has to manage the trading and quote process and has to guarantee the provision of liquidity, when necessary by taking the other side of the market. The data set contains time stamped prices, volumes and bid-ask quotes of the particular transactions. The sample period covers three months from November 1990 till January 1991, corresponding to approximately 60,000 transactions. The preparation of the data set follows along the lines of the work of Engle (2000). All trades before 9:30 a.m. and after 4:00 p.m. and all trades without a reported bid and ask quote are deleted. Furthermore the 11/22/90 (Thanksgiving), the 11/23/90, the 12/23/90 and the 01/01/91 as well as all overnight durations and zero durations are discarded. After this procedure the resulting data set contains 52,540 observations.

In order to generate buy/sell volume durations, the particular trades have to be identified as either buy or sell transactions. The initiation of trades is inferred indirectly from the price and the quote process. The most commonly used methods of inferring the trade direction are the tick test, the quote method as well as hybrid methods which combine both methods (see e.g. Finucane,2000). The tick test uses previous trades to infer the trade direction. According to this method a trade is classified as buy (sell) if the current trade occurs at a higher (lower) price than the previous trade. If the price change between the both transactions is zero, the trade classification is based on the last price that differs from the current price. The quote method is based on the comparison of the transaction price and the midquote. When the price is above (below) the midquote, then the trade is classified as buy (sell). Here, I use a combination of both methods as proposed by Lee and Ready (1991), where the quote method is used to classify all transactions that do not occur at the midquote[12], and the tick test is used to determine transactions where the transaction price equals the midpoint.

Based on this procedure, 56.79% of the observations are identified as buys. Table 1 shows the summary statistics of the time between particular trades, between buys and sells, respectively, as well as of the volume associated with the corresponding transactions. On average, approximately 2 trades per minute are observed with an average volume of

---

[12]Here for 86.49% of all observations.

approximately 2,000 shares per trade. The volume per buy transactions is significantly higher as for sell transactions which might be strongly related to the fact that the price increased from 104$ to approximately 127$ during the sample period.

Table 1: Summary statistics of inter-trade durations, inter-buy durations, inter-sell durations and the corresponding transaction volumes. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91, corresponding to 61 trading days. 52540 trades.

	Durations			Volume		
	All trades	Buys	Sells	All trades	Buys	Sells
Obs	52540	29835	22705	52540	29835	22705
Mean	0.448	0.78	1.03	1934	2360	1373
Std.dev.	0.61	1.23	1.48	6234	7627	3594

Duration data in minutes.

Based on the transaction process, volume durations are generated by systematically thinning the point process. In the following three different types of volume durations are considered:

(i) The time until $v \geq 10,000$ $(20,000)$ shares are absorbed by the market. Such volume durations have an appeal as very general liquidity measures as they measure the traded volume per time, independent of the side of the market.

(ii) The time until $v \geq= 5,000$ $(10,000)$ shares are bought or sold, respectively. Buy (sell) volume durations allow to get deeper insights into the trading activity on the particular market sides and are associated with execution times of unlimited market orders.

(iii) The time until $v \geq 3,000$ $(5,000)$ shares are traded on both sides of the market simultaneously. Based on this criterion, a market is defined as liquid only if a large volume can be traded quickly on both market sides. Thus, this type of volume durations correspond to the waiting time until an unlimited market order on both sides of the market is executed.

Table 2: Summary statistics of the different types of volume durations. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91, corresponding to 61 trading days.

	(i) buy or sell		(ii) buy		(ii) sell		(iii) buy and sell	
shares	10,000	20,000	5,000	10,000	5,000	10,000	3,000	5,000
Obs	6895	3930	7486	4576	3842	2263	3956	2950
Mean	3.39	5.92	3.12	5.08	6.03	10.19	5.88	7.87
Std.dev	3.41	5.47	4.03	5.81	6.73	10.12	5.56	7.21
LB(20)	7459	4898	4828	3711	1115	631	1988	1255

Summary statistics of durations in minutes.

Table 2 presents descriptive statistics of the different types of volume durations. Note that the corresponding aggregation levels are chosen in a way that ensures a satisfying number of observations per trading day. Thus, the mean durations are approximately between 3 and 10 minutes corresponding to approximately 40 and 130 observations per trading day, respectively.
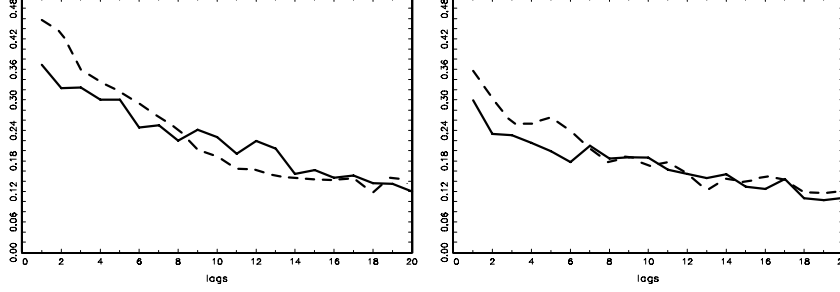
Figure 1: Autocorrelation functions of the time until a volume independent of the market side (left) and a buy volume (right) is traded. Solid lines: 10,000 shares (left), 5,000 shares (right). Broken lines: 20,000 shares (left), 10,000 shares (right).
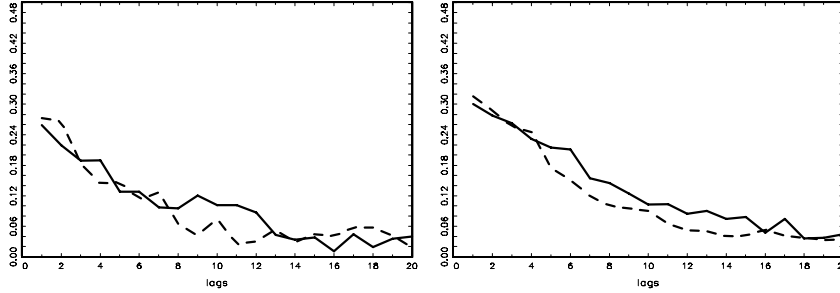
Figure 2: Autocorrelation functions of the time until a sell volume (left) and a volume on both sides of the market (right) is traded. Solid lines: 5,000 shares (left), 3,000 shares (right). Broken lines: 10,000 shares (left), 5,000 shares (right).

The Ljung-Box (LB) statistic formally tests the null hypothesis that the first 20 autocorrelations are zero and is distributed as a $\chi^2(20)$ with a 5% critical value of 31.41. Figures 1 and 2 depict the corresponding autocorrelation functions (ACF). We observe strong serial correlation patterns, thus the null hypothesis is easily rejected for all types of volume durations. In general, the ACF's look very similar, where for volume durations on higher aggregation levels a significantly higher first lag autocorrelation is observed than for waiting times based on lower aggregation levels. Moreover, slightly different patterns of the ACF of the particular duration concepts are observed. The strongest serial dependence is

found for volume durations measured independently of the side of the market. An interesting finding is that sell volumes show a significantly lower serial dependence than buy volumes.
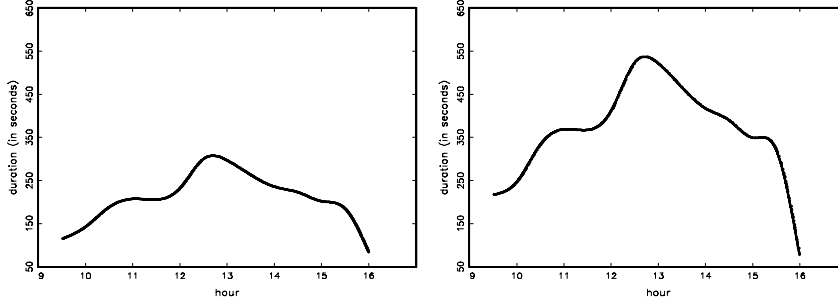
Figure 3: Intraday seasonalities of the time until a volume is traded independent of the side of the market. Nonparametric estimation based on cubic splines (30 minute nodes). Left: $v = 10,000$ shares, right $v = 20,000$ shares.

A well known feature of volume durations (see e.g. Giot, 2000) is a strong impact of intraday seasonality effects. It is assumed that the daily seasonality factor s_t can be approximated by a cubic spline where the nodes are set on each 30 minutes. Thus, I regress the durations on the splines and standardize them by the obtained seasonality components following eq. (13). Figure 3 shows the seasonality patterns of volume durations which do not account for either side of the market. The plot depicts the typical intraday seasonality pattern with high market activities, i.e. small volume durations, in the morning, a significant 'dip' at noon and a relatively active trading before the closure of the market.
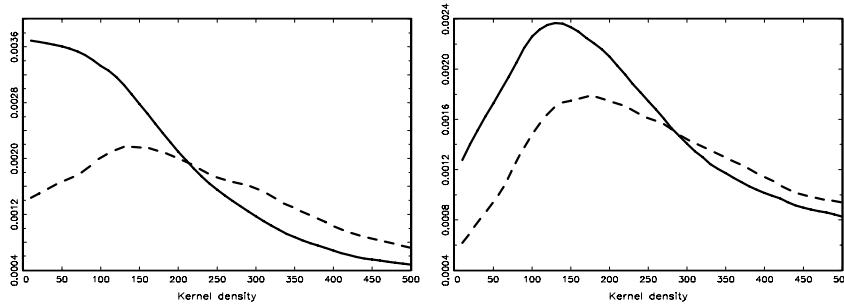
Figure 4: Kernel density estimates of different types of volume durations. Left: Time until a predetermined volume is traded independent of the side of the market. Right: Time until a predetermined volume is traded on each side of the market. Solid line: $v = 10,000$ shares (left), $v = 3,000$ shares (right). Broken line: $v = 20,000$ shares (left), $v = 5,000$ shares (right).

The pictures in figure 4 show kernel density estimates of the seasonal adjusted durations of type (i) and (iii). A well known feature of volume durations is that the density functions of

volume durations become more hump-shaped the higher the aggregation level. This property is one main difference to price durations or inter-trade durations which exhibit density functions that are more similar to the exponential distribution. Thus, modelling of such durations requires the use of ACD models which account for these specific distributional properties.

5. ESTIMATION RESULTS

In the following I, analyze the performance of six different types of ACD models, the basic (linear) ACD specification, the Log-ACD model, the BC-ACD model proposed by Dufour and Engle (2000) as well as the two additional Box-Cox-ACD specifications, BC1-ACD and BC2-ACD. Table 3 shows the regression results of the different models. The estimation is performed by the maximum likelihood procedure of GAUSS. The model selection and comparison are based on the Bayes Information Criterion (BIC) which ascertains, in general, ACD specifications with lag order $p = q = 1$ as the best specification. Note that the time series of the volume durations is re-initialised every trading day, i.e. serial dependencies between observations of different trading days are excluded.

Columns (1)-(5) give the estimates based on the Generalized F distribution. The following conclusions can be drawn: First, all autocorrelation parameters are highly significant. The estimates of α and β indicate that in all regressions the persistence declines for higher aggregation levels. Moreover, again slight differences between the particular duration types are found. Confirming the descriptive statistics, the lowest persistence is observed for sell volume durations, while for volume durations which do not account for the different sides of the market the strongest persistence is found. Diagnostic checks based on the ACD residuals

$$\hat{\epsilon}_t = \frac{\tau_t}{\hat{\bar{\Psi}}_t \hat{s}_t}$$

allow to test whether the used ACD models are appropriate to capture the dynamics and distribution properties of the durations. The Ljung Box (LB) statistics of the $\hat{\epsilon}_t$ time series allow to check whether the residuals are i.i.d. In general, I find a higher reduction of the LB statistics for non-linear conditional mean functions, while the highest reduction is induced by Engle and Dufour's BC-ACD model. The ACD and Log-ACD model, however, do not seem to be appropriate to capture the dynamics of volume durations very well.

Table 3: Estimates of ACD models for different volume durations based on different ACD specifications. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91.

(1): ACD (Generalized F)   (4): BC1-ACD (Generalized F)
(2): Log-ACD (Generalized F)   (5): BC2-ACD (Generalized F)
(3): BC-ACD (Generalized F)   (6): BC2-ACD (Weibull)

	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
	volume measured independently of the side of the market											
	$v = 10,000$ shares						$v = 20,000$ shares					
ω	0.094***	0.047***	0.019***	0.020***	0.017***	0.000	0.120***	0.050***	0.024***	0.027***	0.023***	−0.012**
α	0.303***	0.116***	0.194***	0.178***	0.187***	0.156***	0.397***	0.183***	0.253***	0.234***	0.248***	0.206***
β	0.623***	0.839***	0.950***	0.944***	0.946***	0.953***	0.510***	0.749***	0.936***	0.931***	0.933***	0.935***
δ_1			0.672***	0.685***	0.487***	0.411***			0.598***	0.574***	0.364***	0.388***
δ_2					0.740***	0.694***					0.658***	0.704***
a	2.898***	2.141***	2.680***	2.724***	2.747***	1.261***	2.433***	2.155***	2.222***	2.288***	2,277***	1.529***
m	0.361***	0.517***	0.397***	0.388***	0.386***		0.612***	0.718***	0.686***	0.659***	0.664***	
η	0.695***	0.357***	0.537***	0.539***	0.549***		0.462***	0.334***	0.317***	0.334***	0.330***	
Obs	6895	6895	6895	6895	6895	6895	3930	3930	3930	3930	3930	3930
BIC	−6122	−6107	−6016	−6010	−6011	−6038	−3178	−3145	−3107	−3106	−3107	−3113
MEAN $\hat{\epsilon}_t$	1.000	1.007	1.008	1.003	1.003	1.078	1.000	1.003	1.006	1.003	1.003	1.111
SD $\hat{\epsilon}_t$	0.828	0.813	0.800	0.792	0.792	0.846	0.704	0.694	0.678	0.674	0.673	0.743
LB(20) $\hat{\epsilon}_t$	69.06	120.27	25.43	36.08	31.35	30.43	74.07	89.47	29.18	41.17	32.48	33.68
LB(20) $\hat{\epsilon}_t^2$	44.17	118.47	34.94	46.73	39.71	31.325	86.68	167.36	42.52	60.05	45.43	39.66
pv $\chi^2(2)$	0.000	0.000	0.000	0.000	0.000		0.000	0.008	0.011	0.006	0.006	
	buy volume											
	$v = 5,000$ shares						$v = 10,000$ shares					
ω	0.096***	0.059***	0.027***	0.028***	0.027***	0.026***	0.121***	0.058***	0.029***	0.034***	0.030***	0.011**
α	0.257***	0.091***	0.153***	0.148***	0.151***	0.139***	0.316***	0.115***	0.181***	0.174***	0.180***	0.148***
β	0.670***	0.866***	0.955***	0.947***	0.949***	0.955***	0.593***	0.835***	0.950***	0.942***	0.947***	0.956***
δ_1			0.627***	0.639***	0.520***	0.491***			0.608***	0.556***	0.254*	0.098
δ_2					0.665***	0.674***					0.614***	0.657***
a	1.953***	1.233***	1.623***	1.727***	1.724***	0.941***	2.331***	1.845***	2.155***	2.242***	2.202***	1.109***
m	0.408***	0.718***	0.514***	0.477***	0.478***		0.425***	0.572***	0.474***	0.452***	0.462***	
η	0.550***	0.165	0.339***	0.390***	0.387***		0.701***	0.443	0.567***	0.610***	0.589***	
Obs	7486	7486	7486	7486	7486	7486	4576	4576	4576	4576	4576	4576
BIC	−6843	−6823	−6754	−6747	−6751	−6756	−4174	−4149	−4103	−4104	−4105	−4117
MEAN $\hat{\epsilon}_t$	0.998	1.005	1.007	1.002	1.002	0.974	0.997	1.005	1.007	1.001	1.002	1.042
SD $\hat{\epsilon}_t$	1.141	1.105	1.095	1.085	1.085	1.047	0.989	0.954	0.945	0.942	0.938	0.960
LB(20) $\hat{\epsilon}_t$	54.70	111.29	25.10	28.66	27.50	29.96	75.82	84.34	20.56	28.63	22.55	21.82
LB(20) $\hat{\epsilon}_t^2$	29.64	77.29	14.25	20.29	17.19	15.17	52.77	113.28	26.21	26.73	24.23	14.56
pv $\chi^2(2)$	0.000	0.022	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	
	sell volume											
	$v = 5,000$ shares						$v = 10,000$ shares					
ω	0.256***	0.054***	0.027***	0.026***	0.026***	0.019***	0.297***	0.048***	0.033***	0.030***	0.030***	0.012**
α	0.262***	0.091***	0.143***	0.157***	0.156***	0.126***	0.315***	0.117***	0.167***	0.162***	0.154***	0.138***
β	0.500***	0.826***	0.934***	0.905***	0.904***	0.941***	0.411***	0.821***	0.938***	0.939***	0.938***	0.941***
δ_1			0.568***	0.679***	0.690***	0.538***			0.421***	0.488***	0.690***	0.705***
δ_2					0.679***	0.624***					0.470***	0.487***
a	1.848***	1.472***	1.228***	1.721***	1.730***	1.016***	1.968***	1.449***	1.302***	1.312***	1.314***	1.242***
m	0.450***	0.618***	0.758***	0.499***	0.497***		0.658***	0.853***	0.957***	0.952***	0.952***	
η	0.306***	0.206***	0.038***	0.247*	0.249*		0.224**	0.132***	0.039*	0.046*	0.048	
Obs	3842	3842	3842	3842	3842	3842	2263	2263	2263	2263	2263	2263
BIC	−3647	−3631	−3605	−3600	−3604	−3603	−2081	−2044	−2037	−2034	−2038	−2030
MEAN $\hat{\epsilon}_t$	0.999	1.009	0.986	1.005	1.005	1.009	1.000	1.006	1.006	1.005	0.963	1.074
SD $\hat{\epsilon}_t$	0.966	0.980	0.933	0.958	0.958	0.953	0.828	0.815	0.804	0.803	0.770	0.859
LB(20) $\hat{\epsilon}_t$	109.72	69.51	31.15	38.62	38.85	30.71	83.49	32.31	18.77	18.84	21.09	19.52
LB(20) $\hat{\epsilon}_t^2$	57.82	46.33	24.34	25.27	25.32	22.25	49.14	25.14	17.28	16.36	18.46	16.21
pv $\chi^2(2)$	0.000	0.000	0.003	0.000	0.000		0.022	0.002	0.182	0.383	0.271	

***, **, *: significance on the 1%, 5% or 10% level, respectively.

Table 3 continued: Estimates of ACD models for different volume durations based on different ACD specifications. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91.

(1): ACD (Generalized F) (4): BC1-ACD (Generalized F)
(2): Log-ACD (Generalized F) (5): BC2-ACD (Generalized F)
(3): BC-ACD (Generalized F) (6): BC2-ACD (Weibull)

	volume measured on both sides of the market											
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
	$v = 5,000$ shares						$v = 10,000$ shares					
ω	0.163***	0.042***	0.032***	0.030***	0.032***	0.012***	0.185***	0.040***	0.033***	0.031***	0.032***	0.012**
α	0.287***	0.146***	0.169***	0.172***	0.167***	0.149***	0.292***	0.156***	0.177***	0.179***	0.174***	0.146***
β	0.572***	0.798***	0.942***	0.938***	0.935***	0.937***	0.543***	0.776***	0.930***	0.927***	0.926***	0.940***
δ_1			0.273***	0.356***	0.575***	0.682***			0.205***	0.277***	0.467***	0.516**
δ_2					0.315***	0.324***					0.250***	0.221**
a	1.388***	1.620***	1.594***	1.584***	1.584***	1.362***	1.060***	1.618***	1.594***	1.589***	1.594***	1.432***
m	1.360***	1.048***	1.072***	1.082***	1.080***		2.315***	1.132***	1.156***	1.161***	1.155***	
η	0.286***	0.308***	0.293***	0.288***	0.287***		0.168***	0.284***	0.274***	0.271***	0.272***	
Obs	3956	3956	3956	3956	3956	3956	2950	2950	2950	2950	2950	2950
BIC	−3434	−3363	−3361	−3357	−3361	−3425	−2497	−2446	−2448	−2446	−2450	−2492
MEAN $\hat{\epsilon}_t$	0.993	0.997	0.996	0.997	0.997	1.087	0.988	0.996	0.994	0.997	0.997	1.095
SD $\hat{\epsilon}_t$	0.828	0.791	0.787	0.787	0.786	0.859	0.751	0.740	0.738	0.739	0.739	0.812
LB(20) $\hat{\epsilon}_t$	77.83	15.79	15.32	15.66	15.18	16.04	65.48	21.42	18.93	18.82	19.35	19.26
LB(20) $\hat{\epsilon}_t^2$	23.00	16.33	16.14	16.68	15.91	16.93	46.53	26.92	26.31	26.03	25.83	25.03
pv $\chi^2(2)$	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	

***, **, *: significance on the 1%, 5% or 10% level, respectively.

Second, for all Box-Cox parameters we find values between 0.2 and 0.6. Thus, for almost every regression the linear specification ($\delta = 1$) as well as the logarithmic form ($\delta \to 0$) is rejected. The Box-Cox parameters generally decline for higher aggregation levels, hence this result indicates that logarithmic specifications are more suitable for higher aggregated durations.

Third, a comparison of the different specifications based on the BIC indicates that the best fit is obtained for the particular Box-Cox specifications, especially the BC1-ACD model, eq. (6), while the basic ACD model and the Log-ACD model show a relatively poor fit. In general, the inclusion of the additional Box-Cox parameter δ_2 improves the fit. However, the results seem not to hold for every type of volume durations. It is shown that the different duration types require different functional forms of the conditional mean function as indicated by the estimates of the particular Box-Cox parameters. The lowest Box-Cox parameters, for example, are obtained for volume durations measured on both sides of the market, therefore this specification looks relatively similar to the logarithmic one. This result is confirmed by the fact that especially for this type of volume durations the fit of the Log-ACD model is significantly better than for other duration types and comparable to the performance of the Box-Cox specifications.

Fourth, in most of the specifications the distribution parameters a, m and η^{-1} are highly significant. Note that the Generalized F ACD model nests the Generalized Gamma ACD model if the heterogeneity variance goes to zero, i.e. if $\eta^{-1} \to 0$, and the Weibull ACD model if both $\eta^{-1} \to 0$ and $m = 1$. The χ^2-values in table 3 correspond to the test statistic of a Wald test of the Generalized F distribution against the Weibull distribution. In most of the specifications the data do not support a reduction from the Generalized F form to more simple distributions, as the corresponding tests are rejected. Column (6) presents the estimation results of the BC2-ACD model based on the Weibull distribution[13]. It can be shown that the simpler distribution causes an upward bias of β toward more persistence. Therefore, it seems that the lack of distributional flexibility of the Weibull ACD model causes spurious persistence of the conditional mean function.[14] Furthermore, based on the BIC, for all regressions the fit of the Weibull ACD model is poorer than for the corresponding Generalized F specifications. Contrarily, focussing on the Ljung-Box statistics of the ACD residuals, the difference between both models is less clear. For some regressions the Weibull model seems to capture the inter-duration dynamics even in a better way, leading to higher reductions of the Ljung-Box statistics.

Note that, even though the ACD, BC1-ACD and BC2-ACD model do not ensure the non-negativity of the durations, none of these specifications actually predicted negative durations. Hence, the use of logarithmic forms merely in order to ensure the non-negativity for such applications is not really necessary.

6. Evaluation of the Prediction Performance

A further task of this study is to evaluate the predictive performance of the particular approaches with respect to the different types of volume durations based on out-of-sample forecasts. Therefore, I estimated the models on the basis of the first 56 trading days of the sample while the volume durations of the last 5 days are predicted. One-step forecasts of the duration mean as well as of the duration density are computed. The computation of one-step forecasts of the mean based on ACD models is quite obvious since the estimated conditional expectation of the seasonal adjusted duration of the next spell beginning in t, $\hat{\tilde{\Psi}}_t$, is directly comparable to the actual realization.[15] Thus, in this context the prediction

---

[13]I also estimated the other ACD specifications based on the Weibull distribution and found equivalent results. For ease of exposition only the BC2-ACD model is presented here.

[14]This result is in line with the findings of Lunde (2000) based on evaluations of the Generalized Gamma ACD, Weibull ACD and the Exponential ACD model.

[15]Note that the seasonality adjustment is based on the clock time at the beginning of each spell.

error is given by

$$\frac{\tau_t}{s_t(\tau_{t-1}, \hat{\xi})} - \hat{\tilde{\Psi}}_t.$$

To evaluate the performance of the mean predictions, I use the bias of the forecasts (BIAS), the Root Mean Square Forecast Error (RMSFE), as well as the correlation coefficient (CORR) between the predictions and the corresponding outcomes as loss functions.

To evaluate the density forecasts I apply the method proposed by Diebold, Gunther, and Tay (1998) based on the work of Rosenblatt (1952). The evaluation is based on the probability integral transform

$$z_t = \int_{-\infty}^{\tau_t} f_\tau(s) ds, \tag{15}$$

where f_τ denotes the p.d.f. of τ_t. Diebold, Gunther, and Tay (1998) showed that under the null hypothesis, i.e. correct density forecasts, the distribution of the z_t series is i.i.d. uniform. Hence, testing the z_t series against the uniform distribution allows to evaluate the performance of the density forecasts[16]. Therefore, I categorized the probability integral transforms z_t and computed a χ^2 goodness-of-fit test on the basis of the frequencies of the particular categories

$$\chi^2 = \sum_{j=1}^{K} \frac{(n_j - np_j)^2}{np_j}.$$

K denotes the number of categories (20, in this case), n_j the number of observations in category j and p_j the estimated probability to observe a realization of z_t in category j.

Table 4 reports the prediction results based on the five different Generalized F ACD specifications as well as on the Weibull ACD model. The entries in the last column denote the p-values based on the goodness-of-fit test of the distribution of the z_t series against the $U[0,1]$ distribution.

In general, the best predictability is found for volume durations measured independently of the side of the market. Thus, general movements on the market can be predicted more precisely than the trading activities on the particular sides of the market. For this type of financial durations, the predictive performance with respect to both the mean and the density function rises with the aggregation level.

---

[16]For more details see e.g. Bauwens, Giot, Grammig, and Veredas (2000), who applied this concept to the comparison of alternative financial duration models.

Table 4: One-step out-of-sample forecasts for different types of volume durations based on different ACD specifications. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91, 61 trading days.

(1): ACD (Generalized F)   (4): BC1-ACD (Generalized F)

(2): Log-ACD (Generalized F)   (5): BC2-ACD (Generalized F)

(3): BC-ACD (Generalized F)   (6): BC2-ACD (Weibull)

	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
	volume measured independently of the side of the market											
	$v = 10,000$ shares						$v = 20,000$ shares					
BIAS	0.088	0.092	0.069	0.081	0.076	0.010	0.089	0.083	0.070	0.081	0.075	−0.009
RMSE	0.638	0.625	0.625	0.622	0.624	0.614	0.525	0.526	0.514	0.514	0.513	0.502
CORR	0.201	0.219	0.213	0.225	0.218	0.221	0.274	0.228	0.274	0.280	0.280	0.281
p-val. $\chi^2(19)$	0.002	0.000	0.002	0.000	0.002	0.000	0.527	0.745	0.784	0.418	0.790	0.811
	buy volume											
	$v = 5,000$ shares						$v = 10,000$ shares					
BIAS	0.116	0.073	0.079	0.096	0.094	0.108	0.114	0.091	0.081	0.096	0.087	0.042
RMSE	0.889	0.867	0.869	0.869	0.871	0.871	0.766	0.747	0.745	0.743	0.745	0.737
CORR	0.138	0.138	0.151	0.156	0.152	0.155	0.157	0.151	0.170	0.180	0.172	0.166
p-val. $\chi^2(19)$	0.000	0.000	0.022	0.003	0.010	0.000	0.112	0.000	0.123	0.010	0.084	0.051
	sell volume											
	$v = 5,000$ shares						$v = 10,000$ shares					
BIAS	0.120	0.148	0.104	0.108	0.108	0.096	0.109	0.148	0.095	0.097	0.101	0.055
RMSE	0.694	0.699	0.688	0.687	0.687	0.685	0.560	0.572	0.557	0.557	0.557	0.549
CORR	0.109	0.072	0.103	0.104	0.104	0.094	0.114	0.068	0.089	0.095	0.097	0.095
p-val. $\chi^2(19)$	0.574	0.055	0.260	0.163	0.163	0.403	0.088	0.027	0.032	0.008	0.032	0.203
	volume measured on both sides of the market											
	$v = 3,000$ shares						$v = 5,000$ shares					
BIAS	0.113	0.114	0.101	0.101	0.108	0.037	0.121	0.120	0.110	0.108	0.114	0.041
RMSE	0.565	0.565	0.560	0.560	0.562	0.547	0.548	0.543	0.540	0.539	0.540	0.524
CORR	0.139	0.118	0.126	0.128	0.127	0.134	0.112	0.111	0.116	0.119	0.119	0.126
p-val. $\chi^2(19)$	0.196	0.014	0.000	0.000	0.000	0.000	0.750	0.016	0.049	0.035	0.040	0.009

Furthermore, comparing buy volume durations with sell volume durations I find quite different prediction properties. On the one hand, the predictability of buy volume durations is significantly higher than the predictability of sell volume durations. On the other hand, in contrast to buy volume durations, for sell volume durations the predictive performance declines with the aggregation level. These findings are in accordance with the descriptive statistics (see e.g. figure 2) which indicate quite different ACF's for the various types of volume durations. Hence, the dynamics of trading activities on the sell side seem to be significantly weaker than on the buy side of the market. Economically, this result indicates quite different trading patterns on the particular sides of the market, i.e. different ways in which traders respond to positive or negative price signals. The different properties of buy volume and sell volume durations are also reflected in the waiting times which account for both market sides simultaneously (type (iii)). Here, I find a superposition of the different

effects leading to a prediction performance which is in between the performances of buy volume and sell volume durations.

With respect to mean predictions, the following findings can be summarized: First, the particular Box-Cox specifications as well as, at least for some regressions, the (linear) ACD model have the best predictive performance. In general, the best results are obtained for the BC1-ACD and the BC2-ACD model, where the differences between the particular specifications are relatively small, however. The Log-ACD model is generally a poor forecaster, except for volume durations measured on both sides of the market. Second, the predictive power of the Weibull ACD model is very similar to the performance of the Generalized F model, and even better for some regressions. This result is in line with the findings of Dufour and Engle (2000) who illustrated (by analyzing inter-trade durations) that the choice of the distribution does not significantly affect the out-of-sample performance.

The evaluation of the density forecasts provides a less clear picture: First, no unequivocal ranking of the particular ACD specifications can be stated. In general, the BC2-ACD, the BC-ACD and the basic ACD model have the highest predictive power. Especially for volume durations measured on both sides of the market simultaneously, the ACD model clearly outperforms the other specifications. The Log-ACD model as well as the BC1-ACD model are outperformed for all types of volume durations. These findings illustrate that the choice of the conditional mean function does not seem to have a significant impact on forecasts of the density. Second, surprisingly no clear outperformance of the Generalized F ACD specifications compared to the corresponding Weibull ACD models is observed. In particular, for some duration types even better forecasts based on the Weibull distribution are found. Hence, even though the Generalized F distribution allows for a better fit of the data, it provides no significant improvement of out-of-sample density forecasts.

7. THE IMPACT OF PAST MARKET ACTIVITIES

The goal of this section is to investigate the impact of explanatory variables associated with past trading activities, even when for autoregressive dependencies is controlled. In order to include explanatory variables associated with the beginning of each spell, the ACD model (here the BC-ACD model, eq. (5)), is extended as follows:

(16)

$$((\theta_t \zeta_t)^{\delta_1} - 1)/\delta_1 - x'_{t-1}\gamma = \omega + \sum_{j=1}^{p} \alpha_j (\tau_{t-j}^{\delta_2} - 1)/\delta_2 + \sum_{j=1}^{q} \beta_j (((\theta_t \zeta_t)^{\delta_1} - 1)/\delta_1 - x'_{t-1-j}\gamma).$$

Note that based on this formulation, the explanatory variables enter the model statically, i.e. without an infinite lag structure. Alternatively, one could drop the subtraction of $x'_{t-1-j}\gamma$ in the last term which would correspond to a dynamic form.

From an economic point of view, five questions have to be answered: First, which impact have strong imbalances between the buy and the sell side of the market on the expected (buy or sell) volume duration? Second, which influence has the occurrence of strong price signals, e.g. caused by news announcements? Third, do large changes in absolute prices affect the future market activity? Fourth, are there any asymmetries with respect to the direction of the price changes? I.e., do traders react differently when negative instead of positive price movements have been observed? Fifth, does a high volatility also imply a high volume per time in subsequent trading periods?

In order to answer these questions, I generated five explanatory variables each associated with the trading process of the last 10 minutes before the beginning of the spell:

- RNUMB: The relative number of buy transactions as an indicator for the balance between the two sides of the market.
- RCHBS: The number of relative changes between the buy and the sell side. This variable might be interpreted as an indicator for the 'strength' of a price signal.
- |CDP|: The absolute price change within the last 10 minutes.
- CDP: The (signed) price change within the last 10 minutes.
- CADP: The cumulated absolute price changes from trade to trade within the last 10 minutes as a simple measure for volatility.

Taking into consideration the explanatory variables, I estimated the particular ACD specifications (3)-(7) without finding any significant differences with respect to the estimates of γ. For ease of exposition, in table 5 only the estimation results based on the BC-ACD, eq. (16), are provided.

In general, most of the explanatory variables are found to be significant. Thus they provide additional explanatory power, even though the model controls for serial dependencies. In particular, the following results are summarized: For the variable RNUMB only for sell volume durations a significant coefficient is observed. Interestingly, the coefficient is negative which states a negative relationship between the relative number of buys in the past and the expected sell volume duration. Thus, the more buys have been observed during the last 10 minutes, the higher the expected market activities on the sell side, indicating the existence of some cyclical market behaviour. Second, the less changes between the buy and the sell side have been observed (variable RCHBS), the higher the

probability for the existence of a price signal. Since this measure does not account for the direction of the price signal, it is not surprising that for buy volumes or sell volumes, respectively, no significant results are found. Contrarily, for volume durations defined as the time until on both market sides a given volume is traded, a significant negative coefficient is obtained. Thus, the higher the strength of the price signal associated with a low number of buy-sell changes, the higher the market activities on both sides of the market.

Table 5: Estimates of the BC-ACD model with explanatory variables for different types of volume durations. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91, 61 trading days.

	buy or sell volume		buy volume		sell volume		buy and sell volume	
	$v = 10,000$	$v = 20,000$	$v = 5,000$	$v = 10,000$	$v = 5,000$	$v = 10,000$	$v = 3,000$	$v = 5,000$
ω	0.024^{***}	0.022^{***}	0.020^{**}	0.025^{***}	0.064^{***}	0.032^{***}	0.035^{***}	0.035^{***}
α	0.156^{***}	0.198^{***}	0.102^{***}	0.136^{***}	0.158^{***}	0.139^{***}	0.138^{***}	0.144^{***}
β	0.955^{***}	0.951^{***}	0.973^{***}	0.964^{***}	0.846^{***}	0.934^{***}	0.940^{***}	0.924^{***}
δ	0.684^{***}	0.710^{***}	0.657^{***}	0.674^{***}	0.615^{***}	0.531^{***}	0.379^{***}	0.321^{***}
a	2.544^{***}	1.968^{***}	1.144^{***}	1.901^{***}	1.908^{***}	1.269^{***}	1.574^{***}	1.626^{***}
m	0.429^{***}	0.810^{***}	0.801^{**}	0.563^{***}	0.451^{***}	1.019^{***}	1.127^{***}	1.178^{***}
η^{-1}	0.481^{***}	0.225^{*}	0.101	0.438^{**}	0.362^{**}	0.040	0.290^{***}	0.296^{***}
RNUMB	-0.109^{*}	-0.077	0.018	-0.068	-0.307^{***}	-0.162^{*}	-0.336^{***}	-0.250^{***}
RCHBS	-0.025	0.046	0.008	0.051	0.183	0.144	0.237^{**}	0.293^{***}
$|$CDP$|$	-0.575^{*}	-0.542^{***}	-0.590^{***}	-0.640^{***}	-0.672^{***}	-0.696^{***}	-0.627^{***}	-0.706^{***}
CDP	-0.133^{***}	-0.136^{***}	-0.318^{***}	-0.157^{***}	0.009	-0.055	-0.078^{*}	-0.054
CADP	-0.011^{***}	-0.008	-0.009	-0.013^{**}	-0.026^{***}	-0.016^{**}	-0.025^{***}	-0.022^{***}
Obs	6603	3754	7185	4388	3711	2190	3821	2853
BIC	-5722	-2950	-6419	-3899	-3450	-1952	-3201	-2328
MEAN $\hat{\epsilon}_t$	1.007	1.006	1.008	1.007	1.004	1.006	0.997	0.997
SD $\hat{\epsilon}_t$	0.800	0.673	1.081	0.936	0.965	0.794	0.779	0.735
LB $\hat{\epsilon}_t$	29.54	37.76	24.10	19.35	42.69	17.22	21.62	21.46
LB $\hat{\epsilon}_t^2$	36.42	44.35	44.47	17.52	25.81	14.64	17.96	21.77
pv $\chi^2(2)$	0.000	0.015	0.339	0.000	0.000	0.000	0.000	0.000

***, **, *: significance on the 1%, 5% or 10% level, respectively.

Third, as indicated by a significant negative coefficient of $|$CDP$|$, large price changes within the last 10 minute increase every type of volume durations. Fourth, buy volume durations, as well as volume durations measured independently of the type of the corresponding trades, are significantly negative related to the direction of past price changes. Thus, past positive (negative) price changes decrease (increase) the expected (buy) volume duration. Interestingly, sell volume durations do not seem to be significantly affected by the sign of past price changes. This result might be explained by the existence of some 'asymmetry effects' with respect to the behaviour of traders on the particular sides of the market. Therefore, trading activities on the buy side of the market are stronger related to general price movements than the trading on the sell side. This result is in accordance with the fact that buy volume durations indicate higher serial dependencies than sell volume

durations (see section 6). Fifth, for the relationship between the past volatility and the expected volume per time a significantly negative dependency is observed. Thus, volatile market periods are followed by more active trading phases. In order to analyze whether the inclusion of explanatory variables does also improve the out-of-sample forecast power of the model, table 6 compares the corresponding predictions with and without explanatory variables.

Table 6: Out-of-sample forecasts for different types of volume durations based on the BCACD specification with and without explanatory variables. IBM data based on TORQ database from the NYSE, sample period from 11/01/90 to 01/31/91, 61 trading days.
(1): With explanatory variables.
(2): Without explanatory variables.

	(1)	(2)	(1)	(2)
	buy or sell volume			
	$v = 10,000$ shares		$v = 20,000$ shares	
BIAS	0.069	0.069	0.063	0.070
RMSE	0.612	0.625	0.499	0.514
CORR	0.270	0.213	0.334	0.274
p-val. $\chi^2(19)$	0.077	0.002	0.486	0.745
	buy volume			
	$v = 5,000$ shares		$v = 10,000$ shares	
BIAS	0.064	0.079	0.070	0.081
RMSE	0.851	0.869	0.737	0.745
CORR	0.200	0.151	0.213	0.170
p-val. $\chi^2(19)$	0.000	0.022	0.014	0.123
	sell volume			
	$v = 5,000$ shares		$v = 10,000$ shares	
BIAS	0.124	0.104	0.101	0.095
RMSE	0.688	0.688	0.552	0.557
CORR	0.149	0.103	0.170	0.089
p-val. $\chi^2(19)$	0.226	0.260	0.057	0.032
	buy and sell volume			
	$v = 3,000$ shares		$v = 5,000$ shares	
BIAS	0.108	0.101	0.113	0.110
RMSE	0.561	0.560	0.533	0.540
CORR	0.180	0.126	0.194	0.116
p-val. $\chi^2(19)$	0.001	0.000	0.049	0.049

The results indicate that the inclusion of explanatory variables significantly improves predictions of the mean function while density forecasts do not seem to be significantly affected. Therefore, it is worthwhile not only to flexible specifications but also to account for explanatory factors which are not captured by autoregressive variables.

8. CONCLUSIONS

In this paper, I focussed on the econometric analysis of volume durations, i.e. the time until a predetermined volume is traded on the market. I proposed this concept as a valuable

tool to measure the intraday trading activity, which is strongly related to the liquidity of an asset. By using IBM transaction data from the TORQ database provided by the NYSE, different types of volume durations have been analyzed, in particular, durations based on volumes which do not account for the type of the trade, buy volumes and sell volumes, as well as volumes traded on both sides of the market.

For the econometric modelling of volume durations new, types of ACD models, based on Box-Cox transformations of the conditional mean function, have been proposed. I showed that ACD models based on flexible conditional mean functions provide the best fit of the data as well as the best out-of-sample predictions of the duration mean. Furthermore, it was found that different types of volume durations require different functional forms of the conditional mean function. In particular, for volume durations measured on both sides of the market, estimates of the Box-Cox parameters are obtained, which indicates a conditional mean specification that is relatively similar to a logarithmic one. In order to analyze the role of distributional flexibility, I specified the ACD models based on a Generalized F distribution. The regression results indicated that this higher flexibility leads to a better fit of the data, while it does not significantly improve the predictive power of the model. Comparing the predictability of the different types of volume durations, I found the best results for volume durations which do not account for either side of the market. I.e.,in general it is more difficult to predict liquidity movements on the particular sides of the market. Moreover, I found that buy volume durations are better predictable than sell volume durations which might be interpreted with different trading patterns of the particular sides of the market.

A further objective of the paper was to investigate the impact of past market activities. I included explanatory variables capturing market activities of the last 10 minutes before the beginning of the spell. I illustrated that past imbalances between the buy and the sell side lead to cyclical market behaviour, i.e. the more buys (sells) have been observed, the lower the expected sell (buy) volume duration. Additionally, the less changes there are between the particular market sides, the lower is the expected volume per time measured on both sides of the market. Moreover, the higher the volatility as well as the absolute price change within the last 10 minutes, the higher the trading activity in subsequent trading intervals. Including past signed price changes, it is shown that buy volume durations are significantly negative influenced, while sell volume durations do not seem to be affected by the sign of the past price movement. A further important result is that explanatory variables not only have a significant impact on volume durations, but also improve out-of-sample predictions of the duration mean.

REFERENCES

AGGARWAL, R., AND E. GRUCA (1993): "Intraday Trading Patterns in the Equity Options Markets," *Journal of Financial Research*, 16, 285–297.

AL-SUHAIBANI, M., AND L. KRYZANOWSKI (2000): "An Exploratory Analysis of the Order Book, and Order Flow and Execution on the Saudi Stock Market," *Journal of Banking and Finance*, 24, 1323–1357.

BAUWENS, L., AND P. GIOT (2000): "The Logarithmic ACD Model: An Application to the Bid/Ask Quote Process of two NYSE Stocks," *Annales d'Economie et de Statistique*, 60, 117–149.

BAUWENS, L., P. GIOT, J. GRAMMIG, AND D. VEREDAS (2000): "A Comparison of Financial Duration Models Via Density Forecasts," Discussion paper, CORE.

BESSEMBINDER, H. (2000): "Tick Size, Spreads, and Liquidity: An Analysis of Nasdaq Securities Trading Near Ten Dollars," *Journal of Financial Intermediation*, 9(3), 213–239.

BIAIS, B., P. HILLION, AND C. SPATT (1995): "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *The Journal of Finance*, L, 5, 1655–1689.

BLACK, F. (1971): "Toward a Fully Automated Exchange, Part," *Financial Analysts Journal*, 27, 29–34.

CHAN, L. K. C., AND J. LAKONISHOK (1995): "The Behaviour of Stock Prices around Institutional Trades," *Journal of Finacne*, 4, 1147–1173.

CONROY, R. M., R. S. HARRIS, AND B. A. BENET (1990): "The Effects of Stock Splits on Bid-Ask Spreads," *Journal of Finance*, 45, 1285–1295.

DEMSETZ, H. (1968): "The Cost of Transacting," *Quarterly Journal of Eonomics*, 82, 33–53.

DIEBOLD, F. X., T. A. GUNTHER, AND A. S. TAY (1998): "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

DUFOUR, A., AND R. F. ENGLE (2000): "The ACD Model: Predictability of the Time Between Consecutive Trades," Discussion paper, ISMA Centre, University of Reading.

EASLEY, D., N. M. KIEFER, M. O'HARA, AND J. B. PAPERMAN (1996): "Liquidity, Information and Infrequently Traded Stocks," *Journal of Finance*, 4.

EASLEY, D., AND M. OHARA (1991): "Order Form and Information in Securities Markets," *The Journal of Finance*, 46 (3), 905–927.

ELYASIANI, E., S. HAUSER, AND B. LAUTERBACH (2000): "Market Response to Liquidity Improvements: Evidence from Exchange Listings," *Financial Review*, 35, 1–14.

ENGLE, R. F. (1996): "The Econometrics of Ultra-High Frequency Data," Discussion paper, UCSD.

——— (2000): "The Econometrics of Ultra-High-Frequency Data," *Econometrica*, 68, 1, 1–22.

ENGLE, R. F., AND J. LANGE (1997): "Measuring, Forecasting and Explaining Time Varying Liquidity in the Stock Market," Discussion Paper 6129, NBER.

ENGLE, R. F., AND J. RUSSELL (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

FERNANDES, M., AND J. GRAMMIG (2001): "A Family of Autoregressive Conditional Duration Models," Discussion Paper XX, CORE.

FINUCANE, T. J. (2000): "A Dircet test of Methods for Inferring Trade Direction from Intra-Day Data," *Journal of Financial and Quantitative Analysis*, 35, 553–676.

FLEMING, M., AND E.-M. REMOLONA (1999): "Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information," *Journal of Finance*, 54, 1901–1915.

FOSTER, F. D., AND S. VISWANATHAN (1993): "The Effect of Public Information and Competition on Trading Volune and Price Volatility," *Review of Financial Studies*, 6, 23–56.

GEORGE, T. J., G. KAUL, AND M. NIMALENDRAN (1991): "Spread and its Components: A New Approach," *The Review of Financial Studies*, 4(4), 623–656.

GERHARD, F., AND N. HAUTSCH (2001): "Volatility Estimation on the Basis of Price Intensities," *forthcoming Journal of Empirical Finance*.

GIOT, P. (2000): "Time Transformations, Intraday Data and Volatility Models," Discussion Paper 9944, CORE.

GLOSTEN, L. R. . (1987): "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *The jouranl of finance*, 42, 1293–1307.

GLOSTEN, L. R. (1994): "Is the Electronic Open Limit Order Book Inevitable," *Jjournal of Finance*, 49, 1127–1161.

GLOSTEN, L. R., AND L. E. HARRIS (1988): "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 21, 123–142.

GOURIEROUX, C., J. JASIAK, AND G. LEFOL (1999): "Intra-Day Market Activity," *Journal of Financial Markets*, 2, 193–226.

GRAMMIG, J., AND K.-O. MAURER (2000): "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model," *Econometrics Journal*, 3, 16–38.

GREENE, J., AND S. SMART (1999): "Liquidity Provision and Noise Trading: Evidence from the "Investment Dartboard" Column," *Journal of Finance*, 54, 1885–1899.

HAUTSCH, N. (1999): "Analyzing the Time Between Trades with a Gamma Compounded Hazard Model. An Application to LIFFE Bund Future Transctions," Discussion Paper 99/03, CoFE, Konstanz.

————— (2001): "The Generalized F ACD Model," *Mimeo*.

HUANG, R. D., AND H. R. STOLL (1997): "The Components of the Bid/Ask-Spread: A General Approach," *The review of financial studies*, 10, 995–1034.

KEIM, D. B., AND A. MADHAVAN (1996): "The Upstairs Market for Large-Block Transactions: Analysis and Measurement of Price Effects," *The Review of Financial Studies*, 9, 1–36.

KEYNES, J. (1930): *Treatise on Money*. MacMillan.

KIM, O., AND R. E. VERRECCHIA (1994): "Market Liquidity and Volume around Earnings Announcements," *Journal of Accounting and Economics*, 17, 41–67.

KLUGER, B.-D., AND J. STEPHAN (1997): "Alternative Liquidity Measures and Stock Returns," *Review of Quantitative Finance and Accounting*, 8, 19–36.

KYLE, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53 (6), 1315–1335.

LEE, C. M. C., AND M. J. READY (1991): "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733–746.

LUNDE, A. (2000): "A Generalized Gamma Autoregressive Conditional Duration Model," Discussion paper, Aarlborg University.

PIRRONG, C. (1996): "Market Liquidity and Depth on Computerized and Open Outcry Trading Systems: A Comparison of DTB and LIFFE Bund Contracts," *The Journal of Futures Markets*, 16(5), 519–543.

ROSENBLATT, M. (1952): "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470–472.

VEREDAS, D., J. RODRIGUEZ-POO, AND A. ESPASA (2001): "On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach," Discussion Paper 2001/XX, CORE.

WOOD, R. A., T. H. MCINISH, AND J. K. ORD (1985): "An Investigation of Transactions Data for NYSE Stocks," *Journal of Finance*, 40, 723–741.