

Duration and Order Type Clusters

When Traders React and on which Market Side

Wing Lon NG *

University of Muenster

Institute of Econometrics and Economic Statistics

Am Stadtgraben 9, 48143 Muenster, Germany

e-mail: 05wing@wiwi.uni-muenster.de

June 2004

Abstract

This paper introduces a new bivariate autoregressive conditional framework ($ACD \times ACL$) for modelling the arrival process of buy and sell orders in a limit order book. The model contains two dynamic components to describe the observed clustering of durations and order types: a duration process to capture the time structure, combined with a new “Autoregressive Conditional Logit” model in order to display the traders’ order choice. Both processes are adapted to a common natural filtration and modelled simultaneously. It can be shown that the state of the order book as well as the success and the speed of the matching process have a significant influence on the traders’ decisions when and on which side of the market to submit orders and, thus, affect the market’s liquidity.

Key Words: Ultra high frequency transaction data, limit order book, Market Microstructure, Autoregressive Conditional Duration, Box-Tukey transformation, dynamic logit model.

JEL Classifications: C22, C32, C41.

*The author is very grateful to Mark Trede for his valuable comments and to seminar participants at the University of Muenster and at the Deutsche Statistische Gesellschaft for providing helpful discussion and extensive suggestions. He also wants to express his thanks to the IÖW and the Deutsche Forschungsgemeinschaft for financial support.

1 Introduction

There is a large theoretical and empirical literature on the microstructure of financial markets, boosted by the increased availability of ultra-high frequency transaction data. These time stamped trade-by-trade data are characterized by one main feature, namely the irregularity of time intervals between two consecutive observations. Since durations between transactions often reflect the intensity of trading and thus different degrees of the asset's liquidity, they become important variables explaining the development of intraday returns and volatilities in financial markets. Because the time variable is considered as stochastic, the study of financial econometric models requires alternative methods to the ordinary "fixed-time" series analysis. Based on the seminal work by Engle and Russell (1997, 1998) and Engle (2000), who successfully modelled these specific point processes, many studies have concentrated on the further development of the *Autoregressive Conditional (AC)* framework in order to describe limit order book activities more accurately (see, for example, Bauwens and Giot (2001, 2003), Grammig and Hujer (2001, 2002), Fernandes and Grammig (2003)).

Hence, especially duration (*ACD*) and intensity processes (*ACTI*) are often modified and improved, almost with extensions to multivariate models analyzing the whole order book, not only the trades. Hall, Hautsch and McCulloch (2003), for example, model the *joint* intensity of the bid and ask order arrival process and show in their empirical results that the state of the order book has a significant influence on the order behavior. Although their intensity approach, combined with a logit regression, is convenient for multivariate specifications and time varying covariates, it is far less intuitive and forecasts are computationally burdensome (see also Russell (1999), Bowsher (2003), Bauwens and Hautsch (2003)). In contrast, Engle and Lunde (2003) use two time scales in their analysis. They found out that quotes and trades tend to cluster in time in both a deterministic and stochastic way and thus treat the arrival of trades and succeeding quotes as a bivariate dependent point process. Due to the combination of trade and quote durations, a complicated situation arises, which makes the specification of the dependence between duration pairs very difficult and clearly shows that the *ACD* model as the common duration approach has its limits: In contrast to an intensity approach, one can not take new information into account during the actually lasting waiting time, because one can only condition on the information available at the beginning of each duration. Due to this drawback, duration models are usually not suitable for multivariate specifications, because - in a multidimensional context - different duration processes start and end asynchronously without any common time point to be linked with.

This paper solves this problem by introducing an alternative bivariate modelling framework with an extended *ACD* concept considering *all* points of the whole transaction process (without any distinction between bid-, ask-

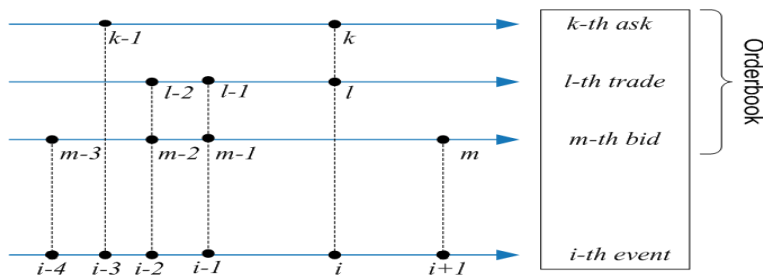


Figure 1: All event points

and trade durations), setting a new point of view in the overall multivariate transaction process, as shown in figure (1). Additionally, to restore the respective type of the order (bid or ask), an innovative dynamic “order type process” is affixed to the *ACD* model. The main objective of this paper is to investigate the arrival process of bid and ask limit orders and to discover what determines the traders’ decisions when and on which side of the market to act. The question addressed here concerns the time varying information set traders refer to before submitting their orders. Comprehending the market conditions under which traders either demand or supply liquidity will lead to a better understanding of the price formation process – the central economic purpose.

Contrary to Engle and Lunde (2003), where two duration processes are modeled jointly, the “mixed” model suggested in this paper is much easier to handle: there is (a) one extended *ACD* model that captures the whole time structure, combined with (b) a new “*Autoregressive Conditional Logit*” model describing the order type with the remaining information of the order book to analyze the traders’ order placement strategy. It is neither a pure duration approach, nor a strict intensity framework. In comparison to other multivariate processes, this bivariate model is easier to understand due to its *AR* structure interlocked twice in the specification.

The outline of this article is as follows: In section 2 the bivariate model is introduced and described. Section 3 shows the smoothing technique and discusses the *ML* estimation. In section 4 the data and the empirical results will be presented, especially with respect to the economic implications. Section 5 concludes.

2 The Model

As high frequency transaction data arrive in irregular time intervals, researchers have to be concerned not only with the variable of genuine interest, but also with the arrival time of each event. In other words, order book data

generally can be characterized by two kinds of random variables. The first one is the time t of the transaction and the other one is the observation \mathbf{Z} (called “marks”, i.e. price, quote, volume) linked with t . Consider the arrival times of events t_0, t_1, t_2, \dots with $t_i \in \mathbb{R}_{\geq 0} \forall i$ as random variables distributed in time by a point process. Usually, it is recorded when traders set bid $(t_k)_{k \in \mathbb{N}}$ and ask $(t_m)_{m \in \mathbb{N}}$ orders and when these orders are filled $(t_l)_{l \in \mathbb{N}}$, yielding a transaction. Thus, for modelling the whole trading development one has to deal with three point processes and their respective arrival times $t_k, t_l, t_m \in \mathbb{R}_{\geq 0} \forall k, l, m$ (see figure (1)). But since transactions are always initiated by either an ask or a bid order, it is sufficient only to record the arrival time and the market side (bid or ask) of incoming orders to get *all* (different) entry points in the whole order book. Hence, $(t_i)_{i \in \mathbb{N}}$ is now the sequence of arrival times of an incoming order, not a transaction. It is to be stressed that the rate of order arrivals only implies the intensity of order submission, but not necessarily fast and frequent trading. It displays first and foremost the order book’s activity, not the market’s liquidity. The purpose of the final bivariate $ACD \times ACL$ model is to describe the order choice and order intensity jointly. An advantage of this new framework is that it does not only allow the prediction of the next order type and its arrival time given the past history, but also easily solves the “zero-duration” problem that frequently occurs: In case of order aggressiveness, for example, a very high demand that cannot be satisfied by one single supplier often must be divided into n transactions to trade with n suppliers. As discernible in figure (2), this one (bid) order has to be matched with several (ask) orders of the opposite market side, leading to split transactions all executed at one time point, probably each one at a different limit price. In univariate models, this “zero-duration” problem is often eluded by aggregating n transactions or, worse, eliminating $n - 1$ transactions. The bivariate model proposed here considers all $n + 1$ orders and their respective arrival time.

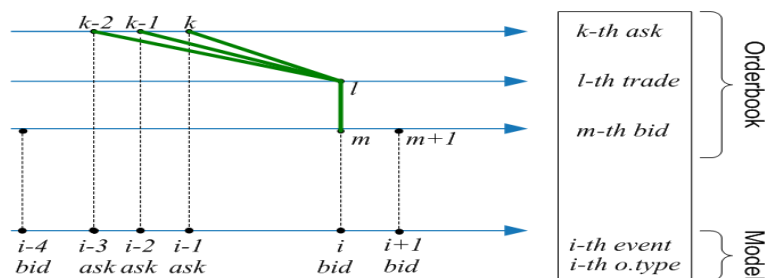


Figure 2: Aggregated transactions

2.1 The ACD Model

Applying a duration framework is the most common approach to describe point processes. Here, the simple *ACD* model as the most popular tool in recent financial econometrics was used to model the elapsing time between consecutive orders and explains duration clustering just through the time dependency of durations. It represents in its simplest form a time series model of time, making it relatively easy to understand, and characterizes a dynamic point process in which the conditional expectation is written as a linear function of past durations.

First of all, introduce a counting function $N(t)$ which simply indicates the number of event arrivals that have occurred before or at time t . This is a monoton-increasing step function with unit increments at each arrival time t . Obviously, $N(t)$ is a simple jump process with $N(t_0) = 0$. Further, define the filtration

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_{i-1} \subseteq \mathcal{F}_i = \sigma(t_0, t_1, \dots, t_{i-1}, t_i)$$

with $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Thus, \mathcal{F}_i is the σ -field generated by all time random variables observed till t_i . The instantaneous probability of an event at t is called the intensity of the process. In time dependent point processes, this intensity is obtained by conditioning on past information. Define the conditional intensity of the process as

$$\begin{aligned} \lambda(t|\mathcal{F}_{N(t)}) &\equiv \lambda(t|N(t), t_1, t_2, \dots, t_{N(t)}) \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) > N(t) | N(t), t_1, t_2, \dots, t_{N(t)})}{\Delta t} . \end{aligned} \quad (1)$$

This function provides a complete description of the point process' full dynamics and is similar to the hazard rate, which is often applied in survival analysis (see Lancaster (1990) or Harrell Jr. (2002)). Now let

$$X_i = t_i - t_{i-1}$$

with $t_0 = 0$, where X_i is the i^{th} duration between the i^{th} and $(i-1)^{th}$ incoming order. Further, define $\Psi_i \equiv E(X_i|\mathcal{F}_{i-1})$ as the conditional duration adapted to the filtration \mathcal{F}_{i-1} , i.e. the conditional expectation of the adjusted duration given the information available at t_{i-1} . Usually, linear ARMA parameterizations are proposed for the conditional mean duration. Therefore, let

$$\begin{aligned} \Psi_i &\equiv E(X_i|\mathcal{F}_{i-1}) \\ &= \Psi(X_{i-1}, \dots, X_1; \theta_1) \\ &= \omega + \alpha_1 X_{i-1} + \dots + \alpha_p X_{i-p} + \beta_1 \Psi_{i-1} + \dots + \beta_q \Psi_{i-q} \\ &= \omega + \sum_{j=1}^p \alpha_j X_{i-j} + \sum_{k=1}^q \beta_k \Psi_{i-k} \quad , \end{aligned} \quad (2)$$

where the parameters $\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ are all included in $\boldsymbol{\theta}_1$. To ensure the stationarity of the process, one must take care of the restrictive constraints for this basic linear specification, in general

$$\begin{aligned} \omega &> 0 \\ \alpha_j, \beta_k &\geq 0 \quad \forall j, k \\ \sum_{j=1}^p \alpha_j + \sum_{k=1}^q \beta_k &< 1 \quad . \end{aligned}$$

This specification clearly implies that the conditional mean adjusts proportionally to recent durations and the influence of these shocks will decline exponentially over time. Since this probabilistic structure of the conditional duration Ψ_i in equation (2) is similar to that of a *GARCH* process, this class of duration models are also called “*autoregressive conditional*”, characterized by the lag lengths p and q of the past (un-/conditional) durations. It is now assumed that

$$X_i = \Psi_i \cdot \varepsilon_i \quad (3)$$

with ε_i as the innovations (see Engle and Russell (1997, 1998)). The essential property here is that the errors $\varepsilon_i \equiv \frac{X_i}{\Psi_i}$, defined as *standardized durations*, are independent and identically distributed random variables, following a special distribution function with the density function $f(\cdot)$, which must be specified:

$$\varepsilon_i \stackrel{i.i.d.}{\sim} f(\varepsilon_i; \boldsymbol{\theta}_1) \quad .$$

Of course, this density must have a non-negative support to avoid negative durations. Further assume that ε_i are independent of X_i . As durations and expected durations are positive, the multiplicative disturbance naturally will have positive probability only for positive values and it must have a mean of unity, i.e.

$$E(\varepsilon_i) = 1 \quad .$$

This requires all temporal dependence of the durations to be captured entirely by the mean function, whereby this hypothesis is testable in practice by using the standardized durations (Engle (2000)).

In the duration analysis literature, the type of this model is referred to as an *Accelerated Failure Time* model which can be viewed as a log-linear regression model for time as well. Equation (3) specifies that the predictor Ψ_i act multiplicatively on the failure time or additively on the log failure time. The effect of Ψ_i is to alter the rate at which a subject proceeds along the time axis, i.e. to accelerate or decelerate the time to “failure” or, here, to submit an order. The crucial assumption for *ACD* models is that the conditional intensity is only specified by the mean equation Ψ_i that is able to summarize the whole dependence structure and is in general defined as

$$\lambda(t|N(t), t_1, t_2, \dots, t_{N(t)}) = \lambda\left(\frac{t - t_{N(t)-1}}{\Psi_{N(t)}}\right) \cdot \frac{1}{\Psi_{N(t)}} \quad ,$$

where

$$\lambda(\cdot) = \frac{f(\cdot)}{1 - F(\cdot)}$$

is the so-called *baseline-hazard*, whose slope depends on the respective distribution function. While Engle (2000) was preferring an Exponential or a Weibull distribution, other authors favoured more flexible alternatives like the Burr- or F-distribution (Fernandes and Grammig (2003), Hautsch (2002b)). Of course, mixture models are also possible (see De Luca and Gallo (2004)). In this paper, a flexible Generalized Gamma distribution (see Johnson, Kotz and Balakrishnan (1994) and Kleiber and Kotz (2003))

$$\begin{aligned} f_{GGamma}(\varepsilon_i | \mathcal{F}_{i-1}; \boldsymbol{\theta}_1) &= f_{GGamma}\left(\frac{X_i}{\Psi_i} | \mathcal{F}_{i-1}; \boldsymbol{\theta}_1\right) \\ &= \frac{\gamma}{X_i \cdot \Gamma(\lambda)} \left(\frac{X_i}{\Psi_i} \cdot \frac{\Gamma\left(\lambda + \frac{1}{\gamma}\right)}{\Gamma(\lambda)}\right)^{\gamma\lambda} \\ &\quad \cdot \exp\left(-\left(\frac{X_i}{\Psi_i} \cdot \frac{\Gamma\left(\lambda + \frac{1}{\gamma}\right)}{\Gamma(\lambda)}\right)^\gamma\right) \end{aligned}$$

was proposed, with λ and γ as the shape parameters of the density function (both also included in $\boldsymbol{\theta}_1$, see figure (5)). $\Gamma(\cdot)$ denotes the incomplete Gamma function. The scope parameter is replaced by $\left(\Gamma\left(\lambda + \frac{1}{\gamma}\right) / \Gamma(\lambda)\right)$, yielding unit expectation. This distribution allows a non-monotonic hazard function, for example, a U-shaped slope for $\lambda\gamma < 1$ and $\gamma > 1$ and vice versa. Clearly, it nests the Gamma distribution for $\gamma = 1$, the Weibull for $\lambda = 1$, the log-normal for $\lambda \rightarrow \infty$, the Exponential for $\gamma = \lambda = 1$, and the $\chi^2(v)$ distribution for $\lambda = \frac{v}{2}$ and $\gamma = 1$ and $\Gamma\left(\lambda + \frac{1}{\gamma}\right) / \Gamma(\lambda) = 2$.

In fact, the assumption of linearity for financial durations is often too restrictive to capture the conditional duration's adjustment process. Hence, several recent models for transaction data have been developed via other dependence structures of the conditional mean to account for nonlinear impacts of past durations (see Bauwens et al. (2003), Hautsch (2004)). Generally, different new types of *ACD* models can be created by varying the functional form $g(\cdot)$ of Ψ_i in the model's equation

$$X_i = g(\Psi_i) \cdot \varepsilon_i \quad .$$

Bauwens and Giot (2001), for example, introduced the *Log-ACD(p,q)* with two possible modifications to explain over and under dispersion and cluster structures as well as *ACD* models, but without any parameter restrictions:

$$\Psi_i = \omega + \sum_{j=1}^p \alpha_j \varphi(\varepsilon_{i-j}) + \sum_{k=1}^q \beta_k \ln(\Psi_{i-k})$$

(see also Bauwens, Galli and Giot (2003)). The new assumption here is that the lagged innovations enters the conditional mean additively, not multiplicatively as considered in the common *ACD* model. Thus, to imply concave news impact curves, define $\varphi(\varepsilon_i) = \ln(\varepsilon_i)$, otherwise just set $\varphi(\varepsilon_i) = \varepsilon_i$ to get convexity. Obviously, this specification allows nonlinear dependency in the conditional mean, but also imposes a very strict adjustment process of the conditional mean to recent durations. Due to the asymptotic convergence to minus infinity at zero of the logarithmic function, one may have an over-adjustment of the conditional mean after very short durations.

Certainly, alternative nonlinear dependence structures are also possible (see, for example, Russell, Tsay and Zhang (2001) or Fernandes and Grammig (2003)), such like the more flexible *Box-Cox-ACD*, originally introduced by Dufour and Engle (2000), modified and extended by Hautsch (2002b)

$$\underbrace{\frac{\Psi_i^{\kappa_1} - 1}{\kappa_1}}_{\Psi_i^{BC}} = \omega + \sum_{j=1}^p \alpha_j \underbrace{\left(\frac{\varepsilon_{i-j}^{\kappa_2} - 1}{\kappa_2} \right)}_{\varepsilon_{i-j}^{BC}} + \sum_{k=1}^q \beta_k \underbrace{\left(\frac{\Psi_{i-k}^{\kappa_1} - 1}{\kappa_1} \right)}_{\Psi_{i-k}^{BC}}, \quad (4)$$

with Ψ_i^{BC} and ε_i^{BC} as the Box-Cox-transformed durations. The Box-Cox version nests both logarithmic *ACD* models and, thus, provide a more accurate description for the adjustment process of the conditional mean to past durations. In contrast to the basic *ACD*, it allows additive innovation shocks (without any interaction with Ψ_i) with $\kappa_1 = \kappa_2 = 1$ as well as nonlinear impact curves for $\kappa_1 < 1$ and $\kappa_2 \leq 1$. For $\kappa_1 \rightarrow 0$ and $\kappa_2 \rightarrow 0$, it would mean that the news impact difference between small innovations ($\varepsilon_i < 1$) is more extensive than between the large ones ($\varepsilon_i > 1$).

Improving the model's fit once more, a new *Box-Tukey-ACD(1,1)* model is proposed in this paper. Instead of the common intercept ω , two new location parameters η_1 and η_2 were introduced to tailor to the data more precisely. Theoretically, it allows even negative observations by shifting the waiting times with the (positive) Box-Tukey parameters η_1 and η_2 and covers the Box-Cox version for $\eta_1 = \eta_2 = 0$:

$$\underbrace{\frac{(\Psi_i + \eta_1)^{\kappa_1} - 1}{\kappa_1}}_{\Psi_i^*} = \sum_{j=1}^p \alpha_j \underbrace{\left(\frac{(\varepsilon_{i-j} + \eta_2)^{\kappa_2} - 1}{\kappa_2} \right)}_{\varepsilon_{i-j}^*} + \sum_{k=1}^q \beta_k \underbrace{\left(\frac{(\Psi_{i-k} + \eta_1)^{\kappa_1} - 1}{\kappa_1} \right)}_{\Psi_{i-k}^*}. \quad (5)$$

Here, $\Psi_i^* = \Psi_i^{BT}$ and $\varepsilon_i^* = \varepsilon_i^{BT}$ are the Box-Tukey-transformed durations. In case of nonlinearity, i.e. the Box-Cox parameter $\kappa_1, \kappa_2 < 1$, stationarity in both cases is ensured for

$$\sum_{k=1}^q \beta_k < 1$$

(see also Russell (1999), Dufour and Engle (2000)).

As illustrated in Engle (2000) and many other studies, the stochastic properties of the trade arrival process and in particular their durations are a decisive reason for volatility. But only examining the trade and its impact on prices and returns is not enough. In fact, most univariate models of ultra-high frequency data is focusing partially on the trades' dimension such as its durations, volumes or price movements, whereas recent studies also demonstrated the importance of the quote's timing and information content (see Engle and Lunde (2003)). While transaction data often only mirror the state of the order book at the intersection of the supply and demand side, quotes allow a deeper insight into the market participants' prior intentions to trade. Therefore, multivariate approaches should be applied to explore the whole electronic limit order books more precisely, taking a closer look at the timing and the content of the bid and ask side. The statistical analysis of the dynamic market process must incorporate not only the distribution of waiting times but also further essential information in the limit order book. It is often neglected that financial electronic markets are constructed for a rapid matching of buyers and sellers of assets. In contrast to former research in market microstructure that is studying preferentially the ("ex-post") consequences, this paper is trying to find the ("ex-ante") reasons for a trader's decision to submit orders.

2.2 The ACL Model

In financial transaction data, a plethora of additional information $\mathbf{Z} = (Z_1, \dots, Z_m)'$ can be observed at the arrival times t (for example, price, volume, quotes, depth, etc.). In this case, the point process $(t_i, \mathbf{Z}_i)_{i \in \mathbb{N}}$ will become "marked", where both the points and the marks are of interest. In general, the common *ACD* model can be extended by including the marks \mathbf{Z}_i in the mean equation for investigating market microstructure hypothesis. Therefore, one just have to add in the market information \mathbf{Z}_i directly into Ψ_i . However, it turns out that the linear specification proposed in Engle and Russell (1997, 1998)

$$\Psi_i = \omega + \sum_{j=1}^p \alpha_j X_{i-j} + \sum_{k=1}^q \beta_k \Psi_{i-k} + \sum_{w=1}^m \tau_w Z_{w,i-1} \quad (6)$$

is insufficient, as the order types (bid or ask) themselves, included in \mathbf{Z}_i , also may depend on the past bid and ask durations and other explanatory variables. To analyze the market participant's preference for a specific order type after observing the limit order book's continuous adjustment process, one needs a separate function $\Upsilon(\cdot)$ that is able to describe his order choice. Regarding a given state of the order book upon arrival, each trader may enter the market as either a buyer or a seller of an asset. Therefore, first

denote

$$\begin{aligned} Y_i &= i^{th} \text{ order type, signaling the market side of the trader} \\ &= \begin{cases} 0 & \text{if order = ask order} \\ 1 & \text{if order = bid order} \end{cases} . \end{aligned}$$

Obviously, $(t_i, Y_i)_{i \in \mathbb{N}}$ is also a marked point process. Since Y_i is a dummy variable representing the market side at t_i , one can model its binary marks by their respective probabilities. Assume that the probability of a bid and an ask order conditioned on \mathbf{Z}_{i-1} is given by

$$P(Y_i = 1 | \mathbf{Z}_{i-1}) \equiv \frac{\exp(\mathbf{Z}_{i-1} \boldsymbol{\tau})}{1 + \exp(\mathbf{Z}_{i-1} \boldsymbol{\tau})} \quad (7)$$

and

$$\begin{aligned} P(Y_i = 0 | \mathbf{Z}_{i-1}) &\equiv 1 - P(Y_i = 1 | \mathbf{Z}_{i-1}) \\ &= \frac{1}{1 + \exp(\mathbf{Z}_{i-1} \boldsymbol{\tau})} \end{aligned} \quad (8)$$

with

$$\begin{aligned} \mathbf{Z}_{i-1} \boldsymbol{\tau} &= \sum_{w=1}^m \tau_w Z_{w,i-1} \\ &= \tau_1 Z_{1,i-1} + \tau_2 Z_{2,i-1} + \dots + \tau_m Z_{m,i-1} . \end{aligned}$$

Here, equation (7) represents the distribution function of the standard logistic distribution that is often used in panel data and microeconometrics. Moreover, from the respective survival function (8) one can derive

$$\begin{aligned} P(Y_i = y_i | \mathbf{Z}_{i-1}) &= P(Y_i = 1 | \mathbf{Z}_{i-1})^{Y_i} \cdot P(Y_i = 0 | \mathbf{Z}_{i-1})^{1-Y_i} \\ &= f_{Logistic}(\mathbf{Z}_{i-1} \boldsymbol{\tau}) \end{aligned} \quad (9)$$

as the density of the logistic distribution which is very important for the model's joint density later on (see Johnson, Kotz and Balakrishnan (1995)). Now define a more comprehensive filtration \mathcal{F}_i^* representing a σ -field that contains the information of all past arrival times and marks till t_i

$$\mathcal{F}_i \subset \mathcal{F}_i^* = \sigma(t_0, t_1, \dots, t_i; Y_0, Y_1, \dots, Y_i; \mathbf{Z}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_i) \quad \forall i .$$

Empirically, traders modify their order placement as soon as the market conditions change. Of course, these strategies depend on the available information when submitting their orders. Monitoring the book and using all new public information, they try to update their information set to optimize further order activities in the whole trading mechanism. Using "static" logit-regressions with lagged variables for modelling this dummy variable

(sell or buy) might be a good choice but not necessary the best. First, considering a time horizon of only one period from t_{i-1} to t_i may be not enough. Indeed, one have to consider that traders do not have a fixed memory limit and are not forced to forget the knowledge they have before t_{i-1} . In fact, they also take the previous information set (in t_{i-2}, t_{i-3}, \dots) into account when making new decisions. Secondly, similar changes of the underlying information set in dissimilar (irregular spaced) time intervals X_i may have different time-varying impacts on the market's dynamics. Formally, one should condition the execution probability on the last state \mathbf{Z}_{i-1} and on the history of the order book. Furthermore, recent studies found out that the probability that an order of a certain type is followed by an order of the same type is relatively high (see figure (7)). It is assumed that the reason for these "order type clusters" often lies in order splitting strategies, imitating behavior, or similar strategic actions by market participants (for example, herding effects, see also Cao, Hansch and Wang (2004) and Foucault, Kadan and Kandel (2003)).

Therefore, to model the dynamic order type process $(Y_i)_{i \in \mathbb{N}}$, the order type probability is conditioned on the natural filtration \mathcal{F}_i^* in a recursive manner, similar to the idea of *GARCH* and *ACD*. To emphasize the analogy of this autoregressive conditional structure with the common *ACD*(p, q) model, just define a general "Autoregressive Conditional Logit" model, *ACL*(u, v), as

$$\begin{aligned}
\Upsilon_i &\equiv P(Y_i = 1 | \mathcal{F}_{i-1}^*) & (10) \\
&= \Upsilon(Y_{i-1}, \dots, Y_1, \mathbf{Z}_{i-1}, \dots, \mathbf{Z}_1; \boldsymbol{\theta}_2) \\
&= \Upsilon(\alpha'_1 Y_{i-1} + \dots + \alpha'_u Y_{i-u} + \beta'_1 \Upsilon_{i-1} + \dots + \beta'_v \Upsilon_{i-v} + \mathbf{Z}_{i-1} \boldsymbol{\tau}) \\
&= \Upsilon\left(\sum_{j=1}^u \alpha'_j Y_{i-j} + \sum_{k=1}^v \beta'_k \Upsilon_{i-k} + \mathbf{Z}_{i-1} \boldsymbol{\tau}\right) \\
&= \left[1 + \exp\left(-\left(\sum_{j=1}^u \alpha'_j Y_{i-j} + \sum_{k=1}^v \beta'_k \Upsilon_{i-k} + \mathbf{Z}_{i-1} \boldsymbol{\tau}\right)\right)\right]^{-1}.
\end{aligned}$$

Alternatively, equation (10) can be rewritten as

$$\ln\left(\frac{\Upsilon_i}{1 - \Upsilon_i}\right) = \sum_{j=1}^u \alpha'_j Y_{i-j} + \sum_{k=1}^v \beta'_k \Upsilon_{i-k} + \mathbf{Z}_{i-1} \boldsymbol{\tau}.$$

It is to be stressed that Υ_i represents the time varying conditional probability $P(Y_i = 1 | \mathcal{F}_{i-1}^*)$ of a bid order which depends on (a) the last state \mathbf{Z}_{i-1} of the order book, (b) the last u order types and finally (c) the last v probabilities, containing previous market information prior to t_{i-1} . For $\beta'_k = 0 \forall k$, the *ACL* model will reduce to the common ("non-autoregressive")

logit model. As implied in the general Autoregressive Conditional framework, this specification suggests that the dynamic conditional probability adjusts to all recent information, but their influence will decay over time. The statistical problem now is to estimate the probability of an order type dynamically, which requires (a) to specify the stochastic process of their arrival times and (b) to estimate all parameters recursively by computing the likelihood function.

To reconstruct the original structure of the order book more accurately, as shown in figure (2), it is recommended to generate certain indicators \mathbf{Z}_i entering the process (10) to reflect the state of the order book at a certain time. Displaying the temporal distance between the order submission of the two market sides, one can introduce the two new interesting duration variables

$$\begin{aligned} Dur_i^{Ask} &= \text{(cumulated) waiting time since the last ask order} \\ &= \begin{cases} t_i - t_{i-1} & \text{if last order is also ask-initiated} \\ Dur_{i-1}^{Ask} + t_i - t_{i-1} & \text{if last order is bid-initiated} \end{cases} \end{aligned}$$

and

$$\begin{aligned} Dur_i^{Bid} &= \text{(cumulated) waiting time since the last bid order} \\ &= \begin{cases} t_i - t_{i-1} & \text{if last order is also bid-initiated} \\ Dur_{i-1}^{Bid} + t_i - t_{i-1} & \text{if last order is ask-initiated} \end{cases} . \end{aligned}$$

Their purpose is to measure the respective order frequency of each market side. Further, introduce an integer variable C_i^{Type} summarizing the number of asks (bids) at time t_i since the last bid (ask):

$$\begin{aligned} C_i^{Bid} &= \begin{cases} 0 & \text{if last order is ask-initiated} \\ C_{i-1}^{Bid} + 1 & \text{if last order is bid-initiated} \end{cases} \\ &= \text{number of new incoming bids since last ask} \end{aligned}$$

and

$$\begin{aligned} C_i^{Ask} &= \begin{cases} 0 & \text{if last order is bid-initiated} \\ C_{i-1}^{Ask} + 1 & \text{if last order is ask-initiated} \end{cases} \\ &= \text{number of new incoming asks since last bid} . \end{aligned}$$

It is obvious that C_i^{Type} is a counting variable, cumulating the number of identical order types clustering on each market side until t_i . In case of order type alteration, this counting variable will be reset to zero for the corresponding side (although there could be more unmatched orders since the last transaction). As a proxy for the buyers' and sellers' order splitting, its aim is to show the bid and ask's cluster length and its switching in the order book.

Since the outstanding volume characterizes the demand and supply side in the limit order book and, thus, displays the market's liquidity, it is usually a good proxy for the execution probability in the particular queues. As it heavily affects the trader's submission behavior, one should include the variables

$$\begin{aligned} QVol_i^{Ask} &= \text{cumulated, unmatched ask volume} \\ QVol_i^{Bid} &= \text{cumulated, unmatched bid volume} \end{aligned}$$

and

$$\begin{aligned} QOrd_i^{Ask} &= \text{number of unmatched ask orders} \\ QOrd_i^{Bid} &= \text{number of unmatched bid orders} \end{aligned}$$

or

$$\begin{aligned} AVol_i^{Ask} &= \frac{QVol_i^{Ask}}{QOrd_i^{Ask}} = \text{average ask volume per order} \\ AVol_i^{Bid} &= \frac{QVol_i^{Bid}}{QOrd_i^{Bid}} = \text{average bid volume per order} \end{aligned}$$

to describe the so-called "order aggressiveness" (high volume orders that induce a large shock and thus move prices). Obviously, $QOrd_i$ and $QVol_i$ represent the thickness of the limit order book and are important factors for describing price effects. In his study of order aggressiveness, Ranaldo (2004) clearly recognizes that "...*patient traders become more aggressive when the own (opposite) side book is thicker (thinner), the spread wider, and the temporary volatility increases.*" In general, order aggressiveness influences the liquidity in an order-driven market essentially, because high-volume orders consume nearly the whole (temporal) liquidity in the market whereas non-aggressive orders provide it.

Furthermore, one has to take into consideration that traders differ by their patience and seek to minimize their trading costs in their order placement strategies. They often weigh between market and limit orders, which is their trade-off between the cost of immediacy and the cost of delayed execution, because a limit order takes time to fill and may fail to fill. A limit order defines a particular price at which the market participant shows his willingness to trade. Certainly, a limit order offers the trader a better price than a market order, but there are costs to submit a limit order and it does not guarantee that it will be executed. Unfilled limit orders are stored in the order book to wait for future execution till being canceled. In contrast, a market order fills immediately at the most attractive price posted by previous limit orders. To find out, whether a price limit was set, denote

$$Lim_i^{Type} = \begin{cases} 0 & \text{if the order was a market order} \\ Price_i^{Type} & \text{if the order was a limit order} \end{cases}$$

with $Type = Ask \text{ or } Bid$

to analyze the limit’s impact on the trader’s order choice due to the influenced waiting times as costs of transactions. Sometimes traders used to “jump” the queue by setting orders with limit prices bettering existing quotes. Of course, one may add further covariates to improve the *ACL* model’s fit (such as the volatility or the spread, etc.) or use other suitable distribution functions.

2.3 The Bivariate Model

To describe the order choice and order intensity jointly, one has to combine the *ACD* model with the *ACL* model. The aim is to model the stochastics of the duration process of all time stamped orders in the order book as well as the dynamics of their order type. The idea is similar to former decomposition approaches, originally proposed by Rydberg and Shepard (2002) to study price movements. In general, these models are two-stage processes, where the first step describes the arrival time (usually the trade duration) and the second explains the marks conditional on the contemporaneous duration. For example, Engle and Russell’s *Autoregressive Conditional Multinomial-ACD* model (2004) investigates price developments by generating a duration process and applying a multinomial distribution for tick movements given the duration. This combination of *ACD* and *ACM* jointly models the distribution of trade durations and possible prices, yielding a price process for transaction data. Likewise, Pohlmeier and Liesenfeld’s *Integer Count Hurdle* model (2003) is splitting the general transaction price process into two components indicating the size and the direction of discrete price changes conditional on the past filtration. But sometimes, we have several different limit prices at the same (transaction) time, as illustrated before (see chapter 2, figure(2)). In this case, one observes prices changes without “temporal changes” which is difficult to explain with decomposition models, because the duration as the basic element is missing here. One can not move to the second stage without passing the first one.

In this paper, the *ACD*×*ACL* models the duration and the order choice simultaneously. The crucial difference here is that both components are not hierarchically structured, allowing a bidirectional dependence structure. Alternatively to the *ACMD* model suggested by Tay et al. (2004), this framework also concentrates on describing the duration and its marks (here, the order type) jointly. Hence, the *ACD*×*ACL* can be written as a system

of equations

$$\Psi_i = \sum_{j=1}^p \alpha_j \varepsilon_{i-j} + \sum_{k=1}^q \beta_k \Psi_{i-k} + \delta_1 \Upsilon_{i-1} \quad (11)$$

and

$$\Upsilon_i = \Upsilon \left(\sum_{j=1}^u \alpha'_j Y_{i-j} + \sum_{k=1}^v \beta'_k \Upsilon_{i-k} + \mathbf{Z}_{i-1} \boldsymbol{\tau} + \delta_2 \Psi_{i-1} \right) \quad (12)$$

(see also the *ACD-GARCH* model suggested by Ghysels and Jasiak (1997)). This *ACD* \times *ACL* model investigates the interesting relationship between the order choice and its duration. One process is indicating the kind of the event (sell or buy), another its time, but none of them is conditioned on the other. It is assumed that both the arrival time and the market side of each order are influenced by the past history of both processes. In particular, the joint distribution of the marked duration process $(X_i, Y_i)_{i \in \mathbb{N}}$ suggested in this framework is modelled as

$$F_{X_i, Y_i}(x_i, y_i) = F_{X_i | Y_i = y_i, \mathcal{F}_{i-1}^*}(x_i) \cdot F_{Y_i | X_i = x_i, \mathcal{F}_{i-1}^*}(y_i) \quad (13)$$

which leads to the interesting mixed density function

$$\begin{aligned} f_{X_i, Y_i}(x_i, y_i) &= f_{X_i | Y_i = y_i, \mathcal{F}_{i-1}^*}(x_i) \cdot f_{Y_i | X_i = x_i, \mathcal{F}_{i-1}^*}(y_i) \\ &= f_{X_i | Y_i = y_i, \mathcal{F}_{i-1}^*}(x_i) \cdot ([\Upsilon_i] \cdot [1 - \Upsilon_i]) \\ &= \underbrace{f_{X_i | Y_i = y_i, \mathcal{F}_{i-1}^*}(x_i)}_{f_{GGamma}} \cdot \underbrace{P(Y_i = y_i | X_i = x_i, \mathcal{F}_{i-1}^*)}_{f_{Logistic}} \quad . \end{aligned} \quad (14)$$

In this bivariate model, the main dependence structure is captured by (11) and (12), each containing and influencing the information for the other process, both adapted to the common filtration \mathcal{F}_{i-1}^* . This specification suggests that the order activity depends on the order decision, and this in turn affects the order frequency again, the history of the book is embedded in both components. Economically, it is assumed that the market behaves asymmetrically if $\delta_1 \neq 0$. For $\delta_1 > 0$, the demand side is acting more frequently and providing more liquidity than the supply side, vice versa for $\delta_1 < 0$. For $\delta_2 \neq 0$, the *ACL* implies an impact of the expected duration on the order submission: if $\delta_2 > 0$, bid spells tend to take more time, again signalling a higher activity on the opposite market side, vice versa for $\delta_2 < 0$. For either $\delta_1 = 0$ or $\delta_2 = 0$, the bivariate process will be reduced to a quasi-decomposed model, where one part represents a subordinated component of the overall transaction process.

3 Estimation and Diagnostics

It is well-known that financial markets pass through hectic periods of increased activity as well as calm slowdowns, reflecting different degrees of

the asset's liquidity. Former studies have found a persisting diurnal pattern of trading activities over the course of a trading day, due to institutional characteristics of organized financial markets (like predetermined opening and closing hours or intraday auctions). For example, in the data set used in this study, the observed waiting times tend to be short in the opening hours of XETRA and NYSE, and tend to be long during lunch time and in the evening hours, as discernible in figure (3). As the rate of information arrival will also vary over the trading day, one has to pay regard to the regular daily seasonality. Therefore, smoothing techniques are required to get deseasonalized observations. Let \tilde{X}_i denote the observed duration. In this paper, a nonlinear kernel regression with an optimal bandwidth h_T^{CV} was performed. After minimizing the Crossvalidation function

$$CV(h_T) = \frac{1}{n} \sum_{i=1}^n \left[\left(\tilde{X}_i - \frac{\sum_{j \neq i} K\left(\frac{t_j - t_i}{h}\right) \tilde{X}_j}{\sum_{j \neq i} K\left(\frac{t_j - t_i}{h}\right)} \right)^2 w(t_i) \right]$$

with $w(\cdot)$ as a nonnegative weighting function and $K(\cdot)$ as the Gaussian kernel function, the diurnal periodic component can be computed by the Nadaraya-Watson estimator

$$\begin{aligned} m(t_i) &= E(\tilde{X}_i | t_i) \\ &= \frac{\sum_{i=1}^n \tilde{x}_i \cdot K\left(\frac{t-t_i}{h_T^{CV}}\right)}{\sum_{i=1}^n K\left(\frac{t-t_i}{h_T^{CV}}\right)} \end{aligned}$$

(see Härdle et al. (2004)). Thus,

$$X_i \equiv \frac{\tilde{X}_i}{m(t_i)}$$

is the deseasonalized duration and should have no diurnal pattern and a mean near unity. To estimate the bivariate model, one must maximize the general likelihood function (see equation (13) and (14)) which is obtained by

$$\begin{aligned} L_{BIV} &= L((x_1, y_1), \dots, (x_n, y_n); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &= \prod_{i=1}^n f_{X_i, Y_i}(x_i, y_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \underbrace{f_{GGamma}(x_i | y_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_1)}_{ACD} \cdot \underbrace{f_{Logistic}(y_i | x_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_2)}_{ACL} \cdot \end{aligned}$$

Taking the logarithm, one gets

$$\begin{aligned}
\mathcal{L}_{BIV} &= \ln L((x_1, y_1), \dots, (x_n, y_n); \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\
&= \sum_{i=1}^n \ln(f_{X_i, Y_i}(x_i, y_i; \boldsymbol{\theta})) \\
&= \sum_{i=1}^n \left[\underbrace{\ln(f_{GGamma}(x_i|y_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_1))}_{ACD} + \underbrace{\ln(f_{Logistic}(y_i|x_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_2))}_{ACL} \right].
\end{aligned}$$

Hence, the likelihood of the *ACD* part generally is

$$\begin{aligned}
L_{ACD} &= L(x_1, \dots, x_n; \boldsymbol{\theta}_1) \\
&= \prod_{i=1}^n f_{GGamma}(x_i|y_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_1) \\
&= \prod_{i=1}^n \frac{\gamma}{x_i \cdot \Gamma(\lambda)} \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma(\lambda + \frac{1}{\gamma})}{\Gamma(\lambda)} \right)^{\gamma\lambda} \cdot \exp \left(- \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma(\lambda + \frac{1}{\gamma})}{\Gamma(\lambda)} \right)^{\gamma} \right),
\end{aligned}$$

from which one can derive the log-likelihood

$$\begin{aligned}
\mathcal{L}_{ACD} &= \ln L(x_1, \dots, x_n; \boldsymbol{\theta}_1) \\
&= \sum_{i=1}^n \ln(f_{GGamma}(x_i|y_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_1)) \\
&= \sum_{i=1}^n \ln \left(\frac{\gamma}{x_i \cdot \Gamma(\lambda)} \right) + \gamma\lambda \cdot \ln \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma(\lambda + \frac{1}{\gamma})}{\Gamma(\lambda)} \right) - \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma(\lambda + \frac{1}{\gamma})}{\Gamma(\lambda)} \right)^{\gamma}.
\end{aligned}$$

To check the *ACD* model's diagnostics, one can examine the properties of the standardized duration such as the mean of unity or their correlation structure. For example, Engle and Russell (1998) and Bauwens and Giot (2001) suggest simply examining the Ljung-Box statistic, although other types of dependence can be investigated, such as nonlinear transformations of the residuals ε_i , for example, squares ε_i^2 or square roots $\sqrt{\varepsilon_i}$. Likewise, theoretical moments suggested by the respective distribution can be compared with the empirical ones. As ε_i is assumed to be Generalized Gamma distributed with λ and γ , the moments of ε_i can be obtained by

$$E(\varepsilon_i^m) = \frac{\Gamma(\lambda)^{m-1} \cdot \Gamma\left(\lambda + \frac{m}{\gamma}\right)}{\Gamma\left(\lambda + \frac{1}{\gamma}\right)}.$$

Another general test is based on the integrated intensity

$$\Lambda_i = \int_{s=t_{i-1}}^{t_i} \lambda(s|\mathcal{F}_{N(s)}) ds \tag{15}$$

over the duration that must be $Exp(\lambda = 1)$ standard exponential distributed under the correct model specification as illustrated in Russell (1999). Providing a transformation of the time scale (from any Non-Poisson process to a homogenous Poisson process), this cumulated hazard function represents - strictly speaking - a Poisson process. The interpretation of this construction is that the original point process does not cease at some point in time which means that further events will definitely occur afterwards. This is realistic for financial point processes, because traders do not stop to submit orders under usual market conditions. Last not least, one can use the probability integral transform

$$\xi_i = \int_{-\infty}^{X_i} f_i(s | \mathcal{F}_{N(s)}) ds \quad (16)$$

as proposed by Bauwens et al. (2003) for density forecast. Here, under correct model specification, ξ_i must be i.i.d. $R(0; 1)$ uniform distributed. Therefore, one can easily derive χ^2 -goodness-of-fit tests to check their uniformity or compute the ACF of ξ_i to analyze their correlation structure. Likewise, simple graphical methods can also be applied.

Further, the likelihood function of the dynamic logit model is given by

$$\begin{aligned} L_{ACL} &= L(y_1, \dots, y_n; \boldsymbol{\theta}_2) \\ &= \prod_{i=1}^n f_{Logistic}(y_i | x_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_2) \\ &= \prod_{i=1}^n [\Upsilon_i]^{y_i} \cdot [1 - \Upsilon_i]^{(1-y_i)} \quad , \end{aligned}$$

or, in the logarithmic form

$$\begin{aligned} \mathcal{L}_{ACL} &= \ln L(y_1, \dots, y_n; \boldsymbol{\theta}_2) \\ &= \sum_{i=1}^n \ln (f_{Logistic}(y_i | \mathbf{z}_i, \mathcal{F}_{i-1}^*; \boldsymbol{\theta}_2)) \\ &= \sum_{i=1}^n y_i \cdot \ln [\Upsilon_i] + (1 - y_i) \cdot \ln [1 - \Upsilon_i] \end{aligned}$$

with

$$\Upsilon_i = \left[1 + \exp \left(- \left(\sum_{j=1}^u \alpha'_j y_{i-j} + \sum_{k=1}^v \beta'_k \Upsilon_{i-k} + \mathbf{z}_{i-1} \boldsymbol{\tau} + \delta_2 \Psi_{i-1} \right) \right) \right]^{-1} .$$

To value the *ACL* model, one can apply various convenient measures that are often used in logit regression, based on modified R^2 -s, or just simply count the correctly forecasted order types in a fourfold-table.

To keep the computational burden acceptable, this paper concentrates on a *Box-Tukey-ACD(1,1)* combined with a simple *ACL(1,1)*. The functions to be maximized jointly in this bivariate $ACD \times ACL$ process are

$$\mathcal{L}_{ACD} = \sum_{i=1}^n \ln \left(\frac{\gamma}{x_i \cdot \Gamma(\lambda)} \right) + \gamma \lambda \cdot \ln \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma(\lambda + \frac{1}{\gamma})}{\Gamma(\lambda)} \right) - \left(\frac{x_i}{\Psi_i} \cdot \frac{\Gamma(\lambda + \frac{1}{\gamma})}{\Gamma(\lambda)} \right)^\gamma$$

with

$$\Psi_i^* = \alpha_1 \varepsilon_{i-1}^* + \beta_1 \Psi_{i-1}^* + \delta_1 \Upsilon_{i-1} \quad (17)$$

and

$$\mathcal{L}_{ACL} = \sum_{i=1}^n y_i \cdot \ln[\Upsilon_i] + (1 - y_i) \cdot \ln[1 - \Upsilon_i]$$

with

$$\Upsilon_i = \Upsilon(\alpha'_1 Y_{i-1} + \beta'_1 \Upsilon_{i-1} + \mathbf{Z}_{i-1} \boldsymbol{\tau} + \delta_2 \Psi_{i-1}) \quad , \quad (18)$$

where

$$\begin{aligned} \mathbf{Z}_{i-1} \boldsymbol{\tau} &= \tau_1 C_{i-1}^{Ask} && + \tau_2 C_{i-1}^{Bid} \\ &+ \tau_3 Dur_{i-1}^{Ask} && + \tau_4 Dur_{i-1}^{Bid} \\ &+ \tau_5 \ln(QVol_{i-1}^{Ask}) && + \tau_6 \ln(QVol_{i-1}^{Bid}) \\ &+ \tau_7 \ln(AVol_{i-1}^{Ask}) && + \tau_8 \ln(AVol_{i-1}^{Bid}) \\ &+ \tau_9 Lim_{i-1}^{Ask} && + \tau_{10} Lim_{i-1}^{Bid} \quad . \end{aligned}$$

4 Dataset and Empirical Results

The transaction data of the Deutsche Telekom stock is extracted from the open order book of the German XETRA system, which is an order-driven market without market makers. The sample includes the whole history of 34419 orders from 21st until 25th August 2000, observed in 5 trading days in 1 week. The continuous trading phase starts after the opening auction at 9 a.m. and ends before the closing auction at 8 p.m. Further, it is interrupted by (at least) two intraday auctions at 1 p.m. and 5 p.m., each lasting at most 120 seconds, as visible in figure (3). The electronic trading is based on an automatic matching algorithm, generally following a strict price-time priority of orders. The ultra-high frequency order book data does not only show the price, the volume and the time stamp (with an accuracy up to one hundredth second) of the transaction but also the initial buy or sell order. Hence, the XETRA data set allows the reconstruction of all submitted (and uncanceled) orders and resulting trades in real time, for it contains the time stamp of all original bid and ask limit orders. The CML-procedure of the Aptech software GAUSS 5.0 was used for the crossvalidation and the joint estimation of the bivariate model proposed above. Descriptive statistics of the durations are listed in table 2. All estimates and their standard errors are reported in table 1 and 4.

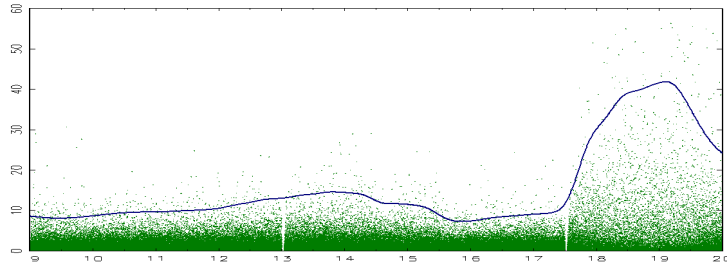


Figure 3: Diurnal pattern of inter-order durations

In the market microstructure theory, the trading and ordering process represents a source of information. Uninformed market participants have to take part to update their pool of information, whereas informed traders only enter the market when they have private information and, therefore, prefer high trading volumes or several orders to capitalize their knowledge before it becomes public. That means, sometimes they are able to exhaust the temporarily existing consumer or producer surplus in the market, due to different price settings. As both informed and uninformed traders are assumed to be indistinguishable upon arrival, this spread between the bid and the ask limit can be viewed as compensation for the risk referred to trading with potentially better informed agents. Informed traders always will make profits to the debit of the uninformed. Other reasons for asymmetric informed market participants are their incapability to study all information in detail at a ultra-high frequency level and, hence, partially ignore them, because supervising and analyzing the whole book will become relatively expensive. If traders are expected to make their decision by using all available information in the order book, one must conclude that different traders have different knowledge due to their different arrival times. Finally, traders often overrate their submission strategy and their own appraisal of the stock's present value.

Knowing the “trends” means reducing the risk. The more information there is in the market, the faster traders have to react. The interesting economic variable here is the time to enter the market. According to the theory, long durations simply suggests that uninformed traders (still) believe that the underlying value of the asset has not changed and only trade because of their own portfolio optimization. Usually, these “noise traders” just want to minimize transaction costs. Contrary, short durations and hence intensive trading signalize the presence of asymmetric information but offer low waiting costs. In other words, long durations signalize a lack of market activities that in turn indicate a longer period of no news, whereas short durations imply the existence of informed trading, where informed traders are assumed to make money by using their informational advantage. Thus,

an understanding of the time varying speed of transactions is important in practice in order to determine *when* to enter the trading platform to demand or supply liquidity. In this study, the *ACD* part (17) of the model detects a strong duration cluster structure (see table (1)).

Table 1: The ACD Model

parameter		coefficient	std. error	t-value
θ_1				
α_1	last standardized duration	0.1178813381	0.0064846351	18.1785615992
β_1	last conditional duration	0.9711609893	0.0025579608	379.6621928857
η_1	Box-Tukey ₁	-0.0717181247	0.0045927347	-15.6155600399
η_2	Box-Tukey ₂	0.3853612046	0.0254903051	15.1179518189
κ_1	Box-Cox ₁	0.4585801408	0.0600686751	7.6342642850
κ_2	Box-Cox ₂	0.5051402760	0.0663805127	7.6097676140
δ_1	P($Y_i = \text{Bid}$)	0.0419168039	0.0051032014	8.2138250705
γ	shape of GGamma distr.	0.6018250970	0.0134601513	44.7116145129
λ	shape of GGamma distr	2.1596758659	0.0855392595	25.2477737067

Table 2: Descriptive statistics of durations

Durations [sec]	\tilde{X}_i	X_i	ε_i
Mean	5.7262382388	1.0890774719	1.0000544571
Median	2.9200000000	0.4835879169	0.6315259036
Variance	104.9368503432	7.4664327905	1.3517846836
Std. dev.	10.2438689148	2.7324774090	1.1626627557
Skewness	8.6430212679	17.0274472369	3.3274679151
Kurtosis	135.0495570623	538.6616915122	28.4053318022

Empirically, we can clearly see (inter-order) duration clusters (in figure (4)), which means that there is a systematic structure in the order book's activity, signalling a certain behavioral pattern of traders. Generally, long durations tend to be followed by long ones and short durations by short ones, $\hat{\alpha}_1, \hat{\beta}_1 > 0$. Because of $\hat{\beta}_1 < 1$, stationarity is ensured, but the duration process shows strong persistence. As both Box-Cox parameters $\hat{\kappa}_1, \hat{\kappa}_2 < 1$, the dependence structure is assumed to be nonlinear, implying concave news impact curves. This would mean that the conditional duration has to be adjusted more extensively during hectic periods (short durations) than in calm spells (long durations). Likewise, both Box-Tukey parameters

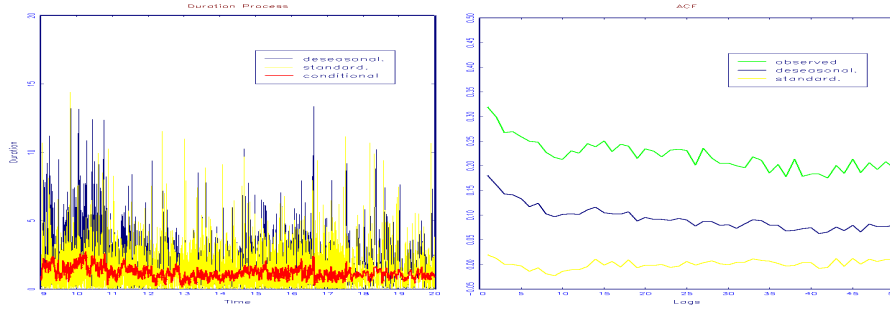


Figure 4: Duration process and ACFs

$\hat{\eta}_1, \hat{\eta}_2 \neq 0$, which indicates that the conditional and standardized durations have to be shifted differently in the overall adjustment process. Standardized duration have to be enlarged with $\hat{\eta}_2$, while the conditional duration must be reduced with $\hat{\eta}_1$. It seems that the demand and the supply side behave asymmetrically in the actual bearish market, as $\hat{\delta}_1 > 0$: bid orders tend to decelerate the process, whereas ask orders accelerate it. The errors $\hat{\varepsilon}_i = x_i / \hat{\Psi}_i$ are Generalized Gamma-distributed with $\hat{\gamma}$ and $\hat{\lambda}$ and seem to have a mean of unity with $\bar{\varepsilon} = 1.0000544571$ (see table 2 and figure (5)). The hazard function is monotonously decreasing at a slow rate. As the Gamma distribution function has no closed form, the integrated intensity was computed by using the exponential formula $\hat{\Lambda}_i = -\ln(1 - F(\hat{\varepsilon}_i))$ and seems to be standard exponential distributed as expected (see figure (6) middle). Finally, the computed probability integral transforms $\hat{\xi}_i$ tend to follow a uniform distribution ($\chi_{10-1}^2 = 0.4657$, see figure (6) left).

Table 3: Comparison of moments

Distribution	Moments	Theoretical	Empirical
$\varepsilon_i \sim GG(\hat{\lambda}; \hat{\gamma})$	$E(\varepsilon_i) =$	1.0000	1.0001
	$Var(\varepsilon_i) =$	1.3687	1.4454
$\Lambda_i \sim Exp(1)$	$E(\Lambda_i) =$	1.0000	1.0001
	$Var(\Lambda_i) =$	1.0000	1.0046
$\xi_i \sim R(0; 1)$	$E(\xi_i) =$	0.5000	0.5008
	$Var(\xi_i) =$	0.8333	0.8380

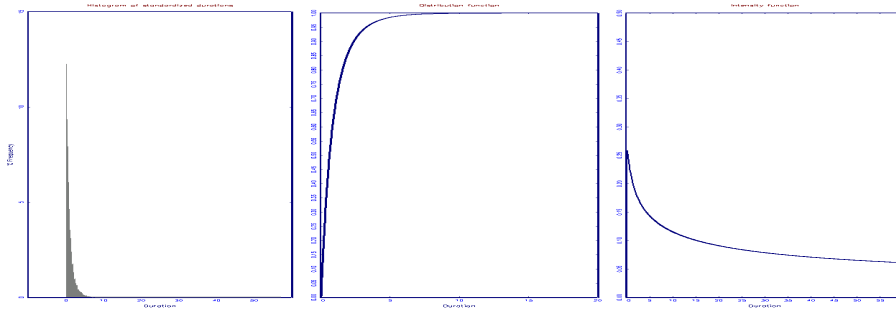


Figure 5: Generalized Gamma distribution of $\hat{\varepsilon}_i$

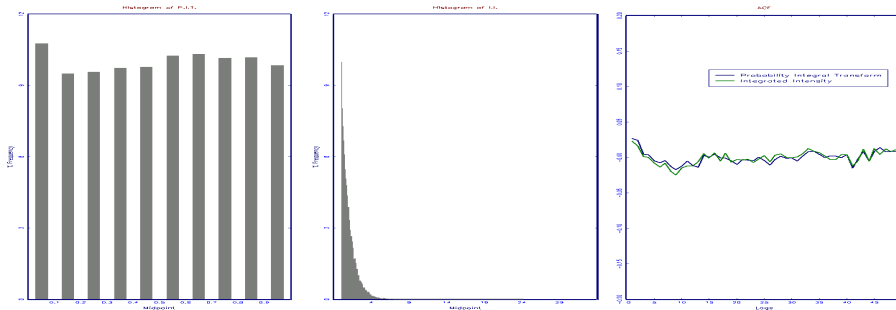


Figure 6: Histogram of $\hat{\xi}_i$ and $\hat{\Lambda}_i$ and their ACFs

As illustrated in the economic theory, the dynamic behavior of durations also influences the time-varying information set of traders. But the *ACD* part of the model only describes the temporal distances between events without distinguishing between different possible kinds of incidents. The characteristics associated with the trading intensity as well as the market's liquidity have a significant impact on the next trader's order choice and his subsequent (more or less) aggressive submission volume. We must have in mind that only different activities of different market participants make trade possible (one sells, one buys). The next step is to investigate the market side from which we observe activities. The new question is not only when but also “*where*” to (re-)act (see table (5)).

According to the theoretical findings, large traded quantities on the particular sides of the market are strong proxies for the existence of information at the current time. It is assumed that informed traders can deduce additional information from the volume that is not noticeable in the price movement. Usually, informed market participants optimize their order placement as the bid-ask spread widens by submitting more limit orders on the opposite market side of the uninformed traders. As illustrated, they try to capitalize

their information as fast as possible before it becomes public in order to make profits at the expense of the uninformed. The higher the differences are, the more volume is absorbed from the particular queue. After the liquidity is taken from the book by one market side, the probability of the occurrence of a trade of the same type in the next instant should decrease.

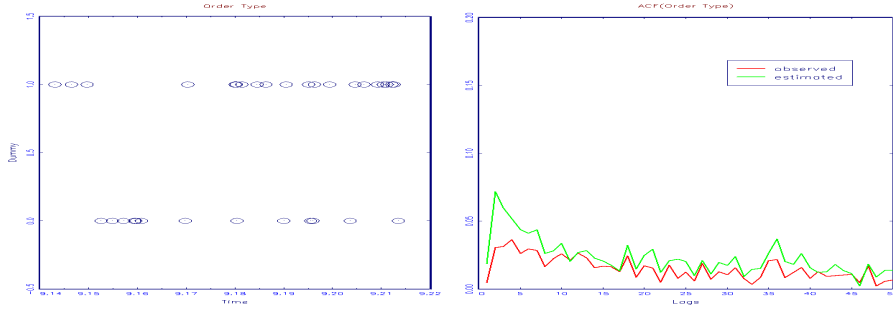


Figure 7: "Order Type Cluster" and ACF of Y_i and \hat{Y}_i

In fact, the data shows that a certain order type (bid or ask) is followed by the same order type sometimes, signalling a certain pattern of repeated order setting (see also Foucault, Kadan and Kandel (2003)). Computing the ACF of the variable "order type" Y_i and its estimated probability \hat{Y}_i , figure (7) show that the dummy is not highly, but always positively auto-correlated. As mentioned above, the reason may lie in order splitting, where traders modify their strategy to maximize the execution probability of all their orders at a minimum cost in order to increase their profitability. Of course, imitating strategies or herding behavior by traders are also possible. Usually, if traders are breaking high-volume orders up into a sequence of smaller orders to execute at a more attractive price overall, these consecutive rows of buys or sells often lead to a sequence of transactions that move the price in the same direction. In this paper, the results of the *ACL* part (18) of the model shows the following order type process (see table (4)):

The probability of a bid order will increase, if the last order type was also a bid, $\hat{\alpha}'_1 > 0$ (vice versa for ask). Secondly, the bid order probability will be the larger, (a) the smaller the preceding bid order probability, $\hat{\beta}'_1 < 0$, (b) the longer the ask sequence's length, $\hat{\tau}_1 > 0$, and (c) the shorter the bid sequence's length, $\hat{\tau}_2 < 0$. Further, a bid order is also more likely, (d) the longer ago the last ask and bid order, $\hat{\tau}_4 > \tau_3 > 0$, and (f) the longer the expected duration, $\hat{\delta}_2 > 0$, again showing asymmetric behavior of both market sides. Likewise, \hat{Y}_i will generally increase, if (g) the log-volume goes down, with $\hat{\tau}_5 < \hat{\tau}_6 < 0$, or if (h) the log-average volume per order increases, whereby $\hat{\tau}_8 > \hat{\tau}_7 > 0$. Finally, the bid order probability will be the larger, (i) the lower the ask limit price, $\hat{\tau}_9 < 0$, and (j) the higher the bid limit price, $\hat{\tau}_{10} > 0$. To value the model's prediction, the fourfold-tables indicate that

Table 4: The ACL Model

parameter		coefficient	std. error	t-value
	θ_2			
α'_1	last order type	0.0984101424	0.0162965929	6.0386943026
β'_1	last order type probability	-0.0183088520	0.0008040544	-22.7706635020
τ_1	ask sequence length	0.0273248626	0.0039147935	6.9798988710
τ_2	bid sequence length	-0.0437771875	0.0094736745	-4.6209300864
τ_3	duration since last ask	0.0030175462	0.0002153646	14.0113371438
τ_4	duration since last bid	0.0098633631	0.0016401193	6.0138082483
τ_5	ln(cumulated ask volume)	-0.1833589062	0.0106635220	-17.1949667628
τ_6	ln(cumulated bid volume)	-0.0299650039	0.0011491461	-26.0758860838
τ_7	ln(average ask volume)	0.1126181585	0.0139881571	8.0509646476
τ_8	ln(average bid volume)	0.2025960135	0.0266302116	7.6077508102
τ_9	ask limit price	-0.0430913882	0.0026728989	-16.1215928576
τ_{10}	bid limit price	0.0698061098	0.0036372368	19.1920716750
δ_2	conditional duration	0.0496884446	0.0083512205	5.9498422775

the dynamic *ACL* model performs a better forecasting than the common logit model (see table 5).

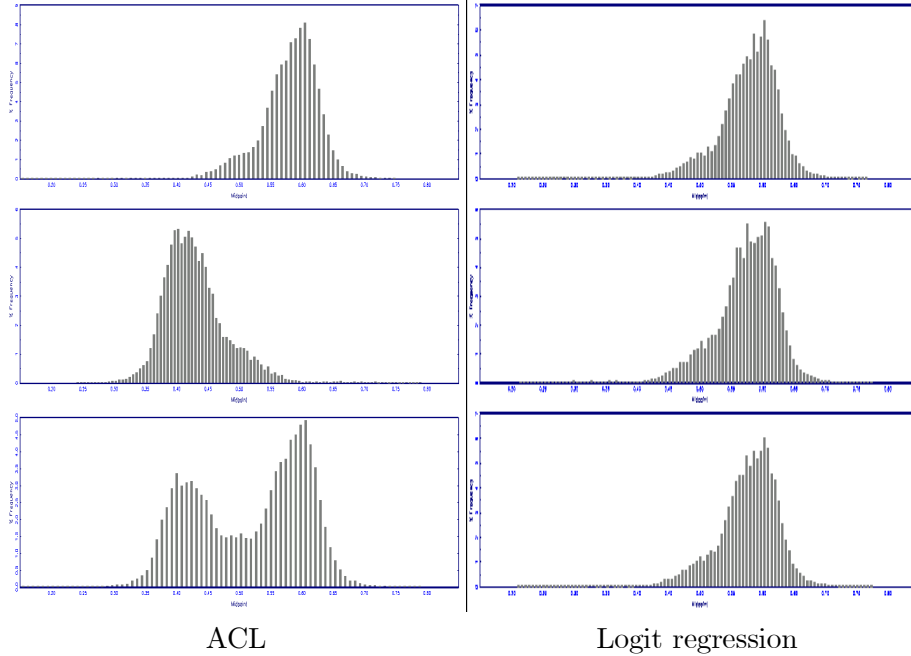


Table 5: The market side

Model	ACD \times ACL		Common Logit		Observed
Order Type [%]	<i>Ask</i> $\hat{Y}_{i \leq 0.5}$	<i>Bid</i> $\hat{Y}_{i > 0.5}$	<i>Ask</i> $\hat{Y}_{i \leq 0.5}$	<i>Bid</i> $\hat{Y}_{i > 0.5}$	Σ
<i>Bid</i> $Y_i=1$	2.7922	54.8653	3.8586	53.7990	57.6576
<i>Ask</i> $Y_i=0$	39.5851	2.7574	3.6958	38.6466	42.3424
Σ	42.3773	57.6227	7.5544	92.4456	100.0000
Correct forecasting	94.4504%		57.4948%		

5 Conclusion

This paper develops a new bivariate modelling framework for analyzing the arrival process of ask and bid orders in an electronic order book market. The econometric approach consists of two parts: On the one hand side, a modified $ACD(1,1)$ model with a flexible Box-Tukey transformation (17) was performed to recover the whole temporal structure of all time stamped events which avoids the occurrence of zero-durations. On the other hand side, a new dynamic $ACL(1,1)$ model (18) was affixed in order to analyze the determinants of the time varying order choice conditioned on the past durations and additional covariates, reflecting the information flow and the market's activity. The main idea was to base both conditional functions on two components jointly, one to model duration clusters, one to describe the dynamic order type with a time dependent probability function, revealing the order book's continuous adjustment process. In contrast to decomposition approaches, both processes are modelled simultaneously which allows bidirectional dependence structures. It is neither a pure duration approach, nor a strict intensity model.

Using detailed transaction data from the German XETRA system, the empirical results show that traders obviously pay strong attention to the order arrival process and the growing queues of the order book before determining when and on which side of the market to act. Facing the available information in the book, they develop their order placement strategies depending on the current thickness of the limit order book and the speed of the overall matching process. The inclusion of the dynamic logit component (10) essentially surpasses the common ACD model (6) and the simple logit regression (7), providing deeper insights into the traders' dynamic order placement strategy and the asymmetric behavior between the supply and demand sides of the market.

References

- [1] **Bauwens, Luc / Giot, Pierre / Grammig, Joachim / Veredas, David (2003):** *A Comparison of Financial Duration Models via Density Forecast*, in: International Journal of Forecasting.
- [2] **Bauwens, Luc / Giot, Pierre (2001):** *Econometric Modelling of Stock Market Intraday Activity*, Kluwer Academic Publishers, Dordrecht, Netherlands.
- [3] **Bauwens, Luc / Giot, Pierre (2003):** *Asymmetric ACD Models: Introducing Price information in ACD Models*, in: Empirical Economics, Volume 28, P. 1-23.
- [4] **Bauwens, Luc / Galli, Fausto / Giot, Pierre (2003):** *The Moments of Log-ACD Models*, Working Paper, CORE & FUNDP Namur.
- [5] **Bauwens, Luc / Hautsch, Nikolaus (2003):** *Stochastic Conditional Intensity Process*, Working Paper, CORE & CoFE.
- [6] **Bowsher, Cliff (2003):** *Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models*, Nuffield Economics Discussion Paper Series, University of Oxford.
- [7] **Campbell, John Y. / Lo, Andrew W. / MacKinley, A. Craig (1997):** *The Econometrics of Financial Markets*, Princeton University Press, Princeton.
- [8] **Cao, Charles / Hansch, Oliver / Wang, Xiaoxin (2004):** *The Informational Content of an Open Limit Order Book*, Discussion Paper, Department of Finance, Pennsylvania State University.
- [9] **Dacorogna, Michel M. / Gencay, Ramazan / Müller, Ulrich / Olsen, Richard B. / Pictet, Olivier V. (2001):** *An Introduction To High-Frequency Finance*, Academic Press, San Diego.
- [10] **De Luca, Giovanni / Gallo, Gampiero M. (2004):** *Mixture Processes for Financial Intradaily Durations*, in: Studies in Nonlinear Dynamics & Econometrics, Vol. 8, Iss. 2.
- [11] **Dufour, Alfonso / Engle, Robert F. (2000):** *The ACD-Model: Predictability of the Time Between Consecutive Trades*, Discussion Papers in Finance, No. 2000-05, University of Reading & University of California.
- [12] **Engle, Robert F. (2000):** *The Econometrics of Ultra-High-Frequency Data*, in: Econometrica, Vol. 68, No. 1, P. 1-22.

- [13] **Engle, Robert F. / Lunde, Asger (2003):** *Trades and Quotes: A Bivariate Process*, in: *Journal of Financial Econometrics*, Vol. 1, No. 2, P. 159-188.
- [14] **Engle, Robert F. / Russell, Jeffrey R. (1997):** *Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model*, in: *Journal of Empirical Finance*, Vol. 4, P. 187-212.
- [15] **Engle, Robert F. / Russell, Jeffrey R. (1998):** *Autoregressive Conditional Duration: A New Model for Irregularly-Spaced Financial Transactions Data*, in: *Econometrica*, Vol. 66, P. 1127-1162.
- [16] **Engle, Robert F. / Russell, Jeffrey R. (2004):** *A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The ACM-ACD Model*, Working Paper, University of Chicago & University of California.
- [17] **Fernandes, Marcelo / Grammig, Joachim (2003):** *A Family of Autoregressive Conditional Durations Models*, in: *Ensaaios Econômicos*, No. 501, ISSN 0104-8910.
- [18] **Foucault, Thierry / Kadan, Ohad / Kandel, Eugene (2003):** *Limit Order Book as a Market for Liquidity*, Working Paper, HEC.
- [19] **Ghysels, Eric / Jasiak, Joanna (1997):** *GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model*, in: *Studies in Non-linear Dynamics & Econometrics*, Vol. 2, Iss. 2.
- [20] **Gourieroux, Christian / Jasiak, John (2001):** *Financial Econometrics*, Princeton University Press, Princeton.
- [21] **Grammig, Joachim / Hujer, Reinhard (2001):** *Forecasting intra-day return volatility using ultra-high-frequency GARCH: does the duration model matter?*, in: *Econometric Studies, Lit, Münster*, P. 203-224.
- [22] **Grammig, Joachim / Hujer, Reinhard (2002):** *Simultaneous modeling of price processes and transaction intensities*, in: *Allgemeines Statistisches Archiv*, Vol. 86, Iss. 1, P. 31-53.
- [23] **Härdle, Wolfgang / Müller, Marlene / Sperlich, Stefan / Werwatz, Axel (2004):** *Nonparametric and Semiparametric Models*, Springer, Heidelberg.
- [24] **Hall, Anthony / Hautsch, Nikolaus / McCulloch, James (2003):** *Estimating the Intensity of Buy and Sell Arrivals in a Limit Order Book Market*, Working Paper, CoFE & UTS.

- [25] **Harrell Jr., Frank E. (2002):** *Regression Modeling Strategies, With Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer, New York et al.
- [26] **Harris, Larry (2003):** *Trading and Exchanges - Market Microstructure for Practitioners*, Oxford University Press, New York.
- [27] **Hautsch, Nikolaus (2002a):** *Testing the Conditional Mean Function of Autoregressive Conditional Duration Models*, Working Paper, CoFE.
- [28] **Hautsch, Nikolaus (2002b):** *Modelling Intraday Trading Activity Using Box-Cox ACD Models*, Working Paper, CoFE.
- [29] **Hautsch, Nikolaus (2004):** *Modelling Irregularly Spaced Financial Data*, Springer, Heidelberg.
- [30] **Hautsch, Nikolaus / Pohlmeier, Winfried (2002):** *Econometric Analysis of Financial Transaction Data: Pitfalls and Opportunities*, in: Allgemeines Statistisches Archiv, Vol. 86, Iss. 1, P. 5-30.
- [31] **Johnson, Norman L. / Kotz, Samuel / Balakrishnan, N. (1994):** *Continuous univariate Distributions*, Vol. 1, 2nd ed., John Wiley & Sons, New York.
- [32] **Johnson, Norman L. / Kotz, Samuel / Balakrishnan, N. (1995):** *Continuous univariate Distributions*, Vol. 2, 2nd ed., John Wiley & Sons, New York.
- [33] **Kleiber, Christian / Kotz, Samuel (2003):** *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, New Jersey.
- [34] **Lancaster, Tony (1990):** *The Econometric Analysis of Transition Data*, Cambridge University Press, New York et al.
- [35] **Liesenfeld, Roman / Pohlmeier, Winfried (2003):** *A Dynamic Integer Count Data Model for Financial Transaction Prices*, Discussion Paper, CoFE.
- [36] **Lunde, Asger (2000):** *A Generalized Gamma Autoregressive Conditional Duration Model*, Discussion Paper, Aalborg University.
- [37] **Rinaldo, Angelo (2004):** *Order aggressiveness in limit order book markets*, in: Journal of Financial Markets, Vol. 7, Iss. 1, P. 53-74.
- [38] **Rydberg, Tina Hviid / Shepard, Neil (2003):** *Dynamics of trade-by-trade price movements: decomposition and models*, in: Journal of Financial Econometrics, Vol. 1, Iss. 1, P. 2-25.

- [39] **Russell, Jeffrey R. (1999):** *Econometric Modelling of Multivariate Irregularly-Spaced High Frequency Data*, Working Paper, University of Chicago.
- [40] **Russell, Jeffrey R. / Tsay, Ruey S. / Zhang, Michael Y. (2001):** *A nonlinear autoregressive conditional model with applications to financial transaction data*, in : *Journal of Econometrics*, Vol. 104, P. 179-207.
- [41] **Tay, Anthony / Ting, Christopher / Tse, Yiu Kuen / Warachka, Mitch (2004):** *Marked Durations and Their Implications for Market Microstructure*, Discussion Paper, Singapore Management University.
- [42] **Tsay, Ruey S. (2002):** *Analysis of Financial Time Series*, John Wiley & Sons, New York.