Comparison of MSACD models

Reinhard Hujer, Stefan Kokot, Sandra Vuletić

University of Frankfurt *

This version: January 10, 2003

Abstract

We propose a new framework for modelling time dependence in duration processes on financial markets. The well known autoregressive conditional duration (ACD) approach introduced by Engle and Russell (1998) will be extended in a way that allows the conditional expectation of the duration process to depend on an unobservable stochastic process which is modelled via a Markov chain. The *Markov switching ACD model* (MSACD) is a very flexible tool for description and forecasting of financial duration processes. In addition, the introduction of an unobservable, discrete valued regime variable can be justified in the light of recent market microstructure theories. In an empirical application we show that the MSACD approach is able to capture several specific characteristics of inter trade durations while alternative ACD models fail.

Keywords: Financial transaction data, autoregressive conditional duration (ACD) models, nonlinear time series models, finite mixture distributions, Markov switching models, EM-algorithm, market microstructure.

JEL classification: C22, C25, C41, G14.

*Institute for Statistics and Econometrics, Faculty of Economics and Business Administration, Johann Wolfgang Goethe-University, Mertonstr. 17, D-60054 Frankfurt. Phone: +49 69 798 28115 Fax: +49 69 798 23673. E-mail: hujer@wiwi.uni-frankfurt.de, kokot@wiwi.uni-frankfurt.de, vuletic@wiwi.uni-frankfurt.de.

Contents

1 **Introduction** **1**

2 **The ACD model** **2**

3 **The Markov switching ACD model** **4**

3.1 Regime switching models in econometrics 4

3.2 The MSACD model 6

3.3 Inference on the latent regime 8

3.4 ML-estimation of the MSACD model 9

3.5 The EM-algorithm for the MSACD model 9

3.6 Statistical inference 11

4 **Empirical application** **14**

4.1 The data set 14

4.2 The flexible scope of the MSACD model 15

4.3 Estimation results 17

5 **Conclusions** **25**

References **26**

1 Introduction

The enormous progress concerning the computer technology makes it possible to collect higher frequency measurements of the economy. Especially, on financial markets it is customary that every single transaction of an asset is recorded electronically with detailed information about the time of occurrence, price and volume and other relevant characteristics. Recently, many of these *ultra high frequency* data sets have become available at relative low cost for academic research. The fullness of theoretical and empirical contributions related to the analysis of *market microstructure* issues are based on transaction data sets.

The introduction of new econometric methods comes along with the development of the relevant theory. One of the most promising new approaches is the autoregressive conditional duration model (ACD) introduced by Engle and Russell (1998) which focuses on the time elapsed between the occurrences of arbitrary trading events. The ACD model combines elements of time series models[1] and econometric tools for the analysis of transition data[2]. Therefore, it is perfectly suited for the analysis of high frequency data sets which unlike most other time series used before in finance and economics are characterized by their *irregularly spacing*. This means that the time between successive observations is not a deterministic constant but rather a random variable itself from which information contents can be exploited. Following the seminal contribution of Engle and Russell (1998), a new branch in the econometric literature quickly emerged that tried to extend their original work in several directions.

Bauwens, Giot, and Grammig (2000) conduct a comparison of the forecast accuracy of various ACD models with respect to a range of duration processes of interest. Despite the resulting variety of competing duration models, until now no satisfactory ACD model in terms of forecast accuracy has been reported that could be used for the prediction of the trading process itself. The main problem is the inability of existing

[1]The ACD approach is related to the GARCH class of models introduced by Engle (1982) and Bollerslev (1986). Many GARCH properties are transferable to ACD models.

[2]For a basic reference see Lancaster (1990).

ACD models to forecast observations in the tails of their distribution, especially very short trade durations, appropriately.

Our intention is to introduce a new reasonable statistical framework for time series of intraday trade durations that can be used for forecasting purposes as well as for tests of the implications of market microstructure models. This will be achieved by the introduction of an additional latent, discrete valued regime variable whose evolution in time is governed by a *Markov chain*. By the way, the inclusion of a latent regime variable in the ACD model can be justified in the light of several recent market microstructure models. The *Markov switching ACD model* (MSACD) provides a very flexible framework which allows to model trade durations resulting from different data generating mechanisms depending on the state of the latent regime and nests many of the existing ACD models as special cases.

This paper is structured as follows: A brief review of the current state of art in ACD modelling will be given in Section 2. Afterwards in Section 3, the MSACD model is introduced and compared to related work on regime switching autoregressive models. Also, we suggest a robust estimation procedure for MSACD models and discuss their applicability and modify test procedures developed by Fernandes and Grammig (2000) and Diebold, Gunther, and Tay (1997) so that they can be applied to MSACD models. In an empirical application in Section 4 we compare the estimation results obtained with the MSACD model to a selection of alternative ACD models. Finally, in Section 5 we summarize our main results and give a perspective on possible issues for future research.

2 The ACD model

The class of autoregressive conditional duration (ACD) models introduced by Engle and Russell (1998) is designed to account for autocorrelation patterns observed in time series of arrival times between successive occurrences of certain events associated with the trading process. The definition of the trading event depends on the specific aim of the study. Examples include the time between successive trades, the time until a price change occurs or until a prespecified number of shares has been traded.[3]

Let $x_n = t_n - t_{n-1}$ be the observed time interval between the $(n-1)^{th}$ and the n^{th} trading event with conditional mean

$$E(x_n|\mathcal{F}_{n-1}) = \psi_n(\mathcal{F}_{n-1};\theta_\psi) \equiv \psi_n \tag{1}$$

which depends on lagged dependent as well as lagged and contemporary exogenous variables, gathered in the filtration $\mathcal{F}_{n-1} = (x_1, \ldots, x_{n-1}, y_1, \ldots, y_{n-1}, y_n)$, and on the corresponding parameter set θ_ψ that determines the conditional mean function. All of the time dependence of the duration process is captured by the conditional mean. The ACD model is defined by some parameterization of this conditional mean and the following decomposition

$$\varepsilon_n = \frac{x_n}{\psi_n} \tag{2}$$

where the stochastic process ε_n is assumed to be i.i.d. with density function $g\left(\cdot;\theta_\varepsilon\right)$ determined by parameters θ_ε and support on the positive real line and an unconditional expectation equal to unity. The choice of $g(\varepsilon_n;\theta_\varepsilon)$ determines the density $f_n\left(x_n \mid \mathcal{F}_n;\theta\right)$ with $\theta = (\theta_\psi, \theta_\varepsilon)$ and will always belong to the same family of distributions as $g(\cdot)$. The flexibility of the ACD model can be altered in at least two ways: by modifying the distributional assumption for the residuals ε_n or the functional form of the conditional mean function ψ_n.

An assortment of admissible distributions arises from the exponential over the Weibull up to the generalized Gamma and Burr distribution respectively. Exponentially distributed random variables are of fundamental importance for modelling waiting times between events from which the families of Weibull, Gamma and generalized Gamma distributions can be derived. Duration data can also be modelled by mixing a specific parametric family of duration distributions with respect to individual

[3] Naturally, the *price* and *volume* duration processes arise from the trade duration series by dropping intervening observations from the sample, thus yielding a 'thinned' or 'weighted' duration process.

heterogeneity. The Burr distribution results as a Gamma mixture of Weibull distributions.[4] Instead of the restrictive exponential and Weibull ACD models originally introduced by Engle and Russell (1998) the use of the generalized Gamma distribution is rather preferred by Lunde (1999) while Grammig and Maurer (2000) propose the Burr distribution for ACD modelling.

In the most simple case an ACD(p, q) model arises when the conditional mean function is determined by a linear autoregressive specification which in analogy to the GARCH model can be transformed into an ARMA $(\max(p, q), p)$ representation from which expressions for the unconditional mean and variance, as well as for the autocorrelation function of the duration process can easily be derived. Alternatively, Bauwens and Giot (2000) suggest a logarithmic specification that closely resembles the EGARCH model of Nelson (1991), implying that the analytical expressions for the unconditional moments of x_n are quite cumbersome in computation as pointed out in Bauwens, Galli, and Giot (2001). Besides the advantage in estimation it allows for more flexibility when additional explanatory variables are included in the model. A transformation of the conditional duration process according to Box and Cox (1964) in addition with an asymmetric shock impact curve which can be justified from empirical findings result in a family of augmented ACD models introduced by Fernandes and Grammig (2001).

3 The Markov switching ACD model

3.1 Regime switching models in econometrics

Apart from the literature on testing for structural changes (e.g. Chow (1960), Goldfeld and Quandt (1965)), models that allow for repeated, discrete changes of regime have been used to model macroeconomic time series with differential behavior in recessions and in expansion phases. In switching regression models, first appeared in

[4] The Burr(μ, κ, σ^2) distribution results when Weibull distributions with random scale parameter $(V \cdot \mu)^{\frac{1}{\kappa}}$ and location parameter κ are mixed according to the Gamma$(\frac{1}{\sigma^2}, \frac{1}{\sigma^2})$ distributed V.

Goldfeld and Quandt (1973), changes in the regime are modelled as the outcome of an unobserved, discrete random variable which identifies the state of the economy in each period. Extensions of this approach lead to models where the regime variable itself is an autoregressive process whose behavior is governed by a hidden Markov chain.

Hamilton (1989) has combined the Markov chain approach for the latent regime with autoregressive dynamics in the observed economic time series. His Markov switching autoregressive model (MSAR) has often been used to model macroeconomic and financial time series (Engel and Hamilton (1990), and Dewachter (2001)). The common link between the MSAR model and the earlier literature on static switching regression models is that both imply that the data generating process of the dependent variable can be described by a discrete mixture density. The regime specific density of the dependent variable is specified to be from some known family of distributions, usually the Gaussian, while the density of the regime variable is left unspecified. The MSAR model has experienced some extensions by allowing for time-varying transition probabilities in the Markov chain as suggested by Filardo (1994) and Gray (1996) or changes in the conditional variances in an ARCH model as proposed by Cai (1994) and Hamilton and Susmel (1994).

Specifically, there are conditional duration approaches which are related to the Markov switching ACD model (MSACD). The treshold ACD model introduced by Zhang, Russell, and Tsay (2001) allows subsamples to have different dynamics for which switchings between them are governed by observables. In order to capture the random flow of nearly unobservable information events on the market Bauwens and Veredas (1999) develop a double stochastic conditional duration model so that in opposite to the origin ACD model expected durations are of random nature as well. Ghysels, Gouriéroux, and Jasiak (1997) propose the stochastic volatility model which accounts for heterogeneity in variances. All these approaches, also including the MSACD framework, imply a mixture distribution model for duration data. The advantage of the flexible MSACD model can be seen in its exquisite ability to reproduce a broad range of different dynamics. It provides dynamics with frequent changes of regimes and likewise it accounts for sudden rare changes.

Starting from the relationship between statistical mixture models for counts of trade events and sequential trade models in the style of Easley, Kiefer, O'Hara, and Paperman (1996) expatiated by **?**, much importance can also be attached to the MSACD model from a theoretical point of view. The MSACD model can be regarded as a generalized sequential trade model which implies that trading evolves in different velocities depending on informational regimes.

3.2 The MSACD model

For financial duration processes, the MSACD framework as a statistical model will be introduced. The general idea is that the conditional mean of the duration time series depends on an unobserved random variable s_n which is regarded as the regime the process is in at time t_n. Formally, the discrete valued stochastic process s_n can assume any value from the set $\mathfrak{J} = \{j \mid 1 \leq j \leq J, J \in \mathbb{N}\}$. In its most general formulation, the MSACD model assumes that the decomposition (2) holds in the sense that $E(\varepsilon_n \mid \mathcal{F}_{n-1}) = 1$. The conditional mean function depends on the unobserved regime variable s_n in the following manner

$$\psi_n = \sum_{j=1}^{J} p\left(s_n = j \mid \mathcal{F}_{n-1}; \theta\right) \cdot \psi_n^{(j)} \tag{3}$$

where $p(s_n = j \mid \mathcal{F}_{n-1}; \theta)$ is the probability that s_n is in state j given the filtration $\mathcal{F}_{n-1}$. The regime specific conditional mean $\psi_n^{(j)} \equiv E(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta)$ depends on an associated set of parameters θ and may have a linear or nonlinear autoregressive specification according to the dynamic of an ordinary ACD model.

The regime variable s_n switches between the states according to a Markov chain which is characterized by a transition matrix P with typical element p_{ji} equal to the transition probability $p_{ji} = p(s_n = j \mid s_{n-1} = i)$. Thus, the state of the process at time t_n depends only on the state of the previous observation.

We assume that the conditional density of the observed duration $f_n(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta)$ depends only on the current regime s_n and on $\mathcal{F}_{n-1}$. Any of the densities we discussed in Section 2 for ordinary ACD models can be used as a conditional density in the MSACD model. Since we cannot observe the realization of the current regime, the relevant density for statistical inference is the *marginal* density $f_n(x_n \mid \mathcal{F}_{n-1}; \theta)$ of the observed duration. In order to evaluate this marginal density in a Markov switching model the *filtered* regime probability

$$\xi^{(j)}_{n+1|n} \equiv p(s_{n+1} = j \mid \mathcal{F}_n; \theta) \tag{4}$$

plays a crucial role. It represents the ex-ante probability for regime j at time t_{n+1} conditional on information available up to time t_n. Filtered regime probabilities can be obtained from a two-step recursion as follows[5]

$$\xi^{(j)}_{n|n} = \frac{\xi^{(j)}_{n|n-1} \cdot f_n\left(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta\right)}{\sum\limits_{k=1}^{J} \xi^{(k)}_{n|n-1} \cdot f_n\left(x_n \mid s_n = k, \mathcal{F}_{n-1}; \theta\right)} \tag{5}$$

$$\xi^{(j)}_{n+1|n} = \sum_{i=1}^{J} p_{ji} \cdot \xi^{(i)}_{n|n}. \tag{6}$$

With a given set of start values $\xi^{(j)}_{1|0}$ and a given parameter vector θ, one can calculate the regime probabilities iteratively. Note, that even though the transition probabilities p_{ji} are constant, the regime probabilities $\xi^{(j)}_{n|n}$ and $\xi^{(j)}_{n+1|n}$ are time-varying. A static mixture model can be regarded as a special case of the Markov switching model. It is based on a restricted transition matrix where the elements of the j-th row are by pairs equal, meaning that $\pi_j \equiv p_{j1} = \ldots = p_{jJ}$ holds. This implies time invariant regime probabilities $\xi^{(j)}_{n+1|n} = \pi_j$ but $\xi^{(j)}_{n|n}$ remains still varying in time.

An issue that has to be addressed, is the specification of the conditional mean function $\psi^{(j)}_n$. There are in principle two possible ways in which lagged forecasts can appear. In the simple case, the current forecast $\psi^{(j)}_n$ is at least a function of $\psi^{(j)}_{n-1}$. Another possible specification is to make $\psi^{(j)}_n$ a function of past forecasts that are

[5] See Hamilton (1994), pp. 692-694 for a proof.

regime independent. However, when regime independent lagged expectations appear in the forecast function, the problem of *path dependence* arises. In this case, the regime dependent expected duration $\psi_n^{(j)}$ depends on the entire sequence of realizations for $(s_1, s_2, \ldots, s_n)$. Since we cannot observe this sequence, we have to consider all J^n possible paths. An evaluation of all possible paths is prohibitively expensive in terms of computational effort even for a moderate sample size . Therefore we apply a heuristic solution based on an aggregation of regime specific conditional means that has been used in the context of Markov switching GARCH models by Gray (1996) and Fong and See (2001). The unconditional expected duration ψ_n is computed by summing over all regime specific conditional expectations $\psi_n^{(j)}$ according to equation (3).

3.3 Inference on the latent regime

Beside the ability to produce forecasts on future durations, in many applications the regimes themselves can be the quantity that the researcher wants to draw inference on. For example, in macroeconomic applications, the regimes can be associated with recession and boom phases in the business cycle. In marketing applications, the inclination to buy certain goods may be related to unobserved heterogeneity among a sample of consumers. Analogously, in financial applications estimates of the regime variable s_n may provide evidence on the presence of agents with private information.

In principle, the regime probabilities given in equation (5) could be employed. A superior inference on the state of the regime may however be obtained by ex-post use of the full sample information. This will provide us with *smoothed inferences* $\xi_{n|N}^{(j)} = p(s_n = j \mid x_N, \mathcal{F}_{N-1}; \theta)$. These may be evaluated using the algorithm of Kim (1994) which consists of a backward recursion starting with the filtered inference $\xi_{N|N}^{(j)}$ obtained from (5) and progressing according to

$$\xi_{n|N}^{(j)} = \xi_{n|n}^{(j)} \cdot \sum_{k=1}^{J} \frac{p_{kj} \cdot \xi_{n+1|N}^{(k)}}{\xi_{n+1|n}^{(k)}}. \tag{7}$$

Application of this algorithm is valid only when s_n follows a first-order Markov chain and when the conditional density of x_n depends only on the current state s_n and on the filtration $\mathcal{F}_{n-1}$.

3.4 ML-estimation of the MSACD model

In the case of regime switching models there are several ways in which estimates of θ may be obtained. The usual approach maximizes the likelihood function based on the marginal density of x_n which is also known as maximizing the *incomplete* likelihood $\mathcal{L}_I(\theta)$, since this likelihood is based on observable quantities only while realizations of the regime variable are unobservable. Thus we estimate θ with an incomplete data set. The log-likelihood function $\ln \mathcal{L}_I(\theta)$ for the MSACD model

$$\ln \mathcal{L}_I(\theta) = \sum_{n=1}^{N} \ln \left[f_n(x_n \mid \mathcal{F}_{n-1}; \theta) \right] \tag{8}$$

has to be maximized numerically under the linear constraints $\sum_{k=1}^{J} p_{kj} = 1$ for all $j \in \{1, ..., J\}$ and additional restrictions for nonnegativity, stationarity[6] and eventually for distributional parameters.

The likelihood function for switching models may have more than one local maximum and these may be located in boundary regions of the parameter space. It is well known that standard maximization algorithms such as the Newton-Raphson may fail or produce nonsensical estimates. In such cases the maximization procedure may be started anew with different start values. It is recommended that estimation should always be repeated several times with different start values in order to make sure that a global maximum has been found.

3.5 The EM-algorithm for the MSACD model

An alternative way of obtaining ML-estimates for Markov Switching models is based on the Expectation-Maximization (EM) algorithm introduced by Dempster, Laird, and

[6] Local stationarity which follows from classical stationarity constraints within each regime is not necessary to guarantee the global stationarity. As shown by Francq and Zakoian (2001), the existence of explosive regimes does not conflict with strict stationarity.

Rubin (1977). Its numerical robustness offers an advantage over standard maximization methods. The basis for the EM-algorithm is the hypothetical situation where we can observe the realization of the sequence of regimes. Defining the random variables $z_n^{(j)} = 1$ if $s_n = j$ and $z_n^{(ji)} = 1$ if $s_n = j$ and $s_{n-1} = i$ and zero otherwise, the *complete* log-likelihood function $\ln \mathcal{L}_C(\theta)$ is given by[7]

$$(9) \quad \ln \mathcal{L}_C(\theta) \quad = \quad \sum_{n=1}^{N} \sum_{j=1}^{J} z_n^{(j)} \cdot \ln[f_n(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta)] + \sum_{n=2}^{N} \sum_{j,i=1}^{J} z_n^{(ji)} \cdot \ln[p_{ji}].$$

The evaluation of the *expected complete* log-likelihood function $\ln \mathcal{L}_{EC}(\theta, \theta_0) \equiv E[\ln \mathcal{L}_C(\theta) \mid \mathcal{X}_N; \theta_0]$ constitutes the first part of the EM-algorithm and is commonly referred to as the *E-step*. In the E-step latent variables, in our case the realizations of the regime indicators, are replaced by their expectations conditional on the observed sample data $\mathcal{X}_N = (x_1, \ldots, x_N, y_1, \ldots, y_N)$ and evaluated using an arbitrary guess for the parameter vector θ_0. The associated *M-step* consists of maximizing $\ln \mathcal{L}_{EC}(\theta, \theta_0)$ with respect to the parameter vector θ, yielding an updated guess for the parameter vector $\hat{\theta}_1$. The same restrictions as in the case of the incomplete log-likelihood function have to be imposed. By repeating these two steps until a prespecified convergence criterion[8] is fulfilled the ML-estimates are found. It can be shown that the final estimates $\hat{\theta}$ maximize both the expected complete log likelihood function as well as the incomplete log likelihood function.[9]

Application of the EM-algorithm has the advantage that the maximization of $\ln \mathcal{L}_{EC}(\theta, \theta_0)$ with respect to the parameters of the ACD model and the transition probabilities can be conducted separately if the equality $\frac{\partial f_n(x_n|s_n=j,\mathcal{F}_{n-1};\theta)}{\partial p_{mk}} = 0, \ \forall \ j, m, k \in (1, \ldots J)$ is satisfied[10]. The first order conditions lead to the estimator for the tran-

[7]The likelihood contribution of the initial state of the regime s_1 can be included in the set of parameters to be estimated. However, it is more convenient to work with a conditional likelihood function, taking the state of the first observation as given.

[8]Often it is suggested to stop the the iteration as soon as $\mid \hat{\theta}_{i+1} - \hat{\theta}_i \mid \leq \delta$ is reached, with δ as a very small number. The EM-algorithm converges slow so that many iterations are necessary to obtain the parameter estimate. Therefore, Aitken acceleration device presented in Böhning, Dietz, Schaub, Schlattman, and Lindsay (1994) can be used.

[9]See Hamilton (1990) for a proof.

[10]This is the case when the regime specific mean function $\psi_n^{(j)}$ is not dependent on past forecasts ψ_{n-1} that are regime independent.

sition probabilities[11] which is essentially equal to the estimator for p_{ji} that we would obtain if the regime variables s_n were observable (i.e. the frequency of observing a transition from state i to state j relative to the frequency of observing state i), again with unobserved quantities replaced by appropriate probabilistic inferences.

3.6 Statistical inference

When conducting specification tests in static mixture and Markov switching models, some care has to be exercised in order to avoid incorrect decisions as a result of the nonstandard distributions of the test statistics involved. An example is testing whether a given data set may be described by a N-regime model or whether $(N-1)$ regimes are sufficient. As argued by Böhning, Dietz, Schaub, Schlattman, and Lindsay (1994) the corresponding likelihood ratio statistic will not have the usual χ^2-distribution, but differ from it substantially even in large samples. Another example is the usual t-statistic for $H_0 : p_{ji} = 0$ against $H_A : p_{ji} > 0$. Under the null hypothesis the transition probability p_{ji} lies on the boundary of the admissible parameter space, thus violating one of the regularity conditions needed in order to derive the asymptotic normal distribution for the t-statistic.

On the other hand, when the number of regimes is *known*, the maximum likelihood estimate of the parameter vector θ has asymptotically a normal distribution with covariance matrix derived from the usual estimates of the information matrix. Hypothesis tests may be conducted in the usual fashion, as long as non of the maintained hypothesis violates the regularity conditions. Therefore, t-statistics for testing whether a particular regression parameter β_{jk} is significantly different from zero may be compared to tabulated critical values of the t-distribution.

Fernandes and Grammig (2000) have introduced a specification test for ordinary ACD models which are based on the discrepancy between the observed and the theoretical density function of the residuals and are, with minor refinements, applicable to the

[11] See Hamilton (1989).

MSACD model as well. In ordinary ACD models the test statistic is easily calculated by noting that the residuals ε_n are independently identically distributed. In contrast to ordinary ACD models the MSACD assumes that residuals follow a known mixture distribution with mean equal to one and time varying higher moments. Therefore, the null hypothesis is

$$H_0 : \exists \quad \theta \in \Theta \quad \text{such that} \quad g(\varepsilon;\theta) \equiv \frac{1}{N} \cdot \sum_{n=1}^{N} g_n(\varepsilon \mid \mathcal{F}_{n-1};\hat{\theta}) = g(\varepsilon) \tag{10}$$

where $g(\varepsilon)$ is the true but unknown density of the residuals and $g(\varepsilon;\theta)$ is the density implied by the parametric MSACD model. In order to make this test operational, a kernel density estimate $\hat{g}(\hat{\varepsilon})$ of the density of the estimated residuals is used and the theoretical density is calculated based on the estimated parameter vector. Thus the observed mean squared distance D_g between the two densities is given by

$$D_g = \frac{1}{N} \sum_{n=1}^{N} \left[g(\hat{\varepsilon}_n;\hat{\theta}) - \hat{g}(\hat{\varepsilon}_n) \right]^2 . \tag{11}$$

Under the null hypothesis (10) the statistic FG has asymptotically a standard normal distribution. FG is given by

$$FG_\varepsilon = \frac{N \cdot h^{0.5} \cdot D_g - h^{-0.5} \cdot \hat{E}_{D_g}}{\sqrt{\hat{V}_{D_g}}}, \tag{12}$$

where h is the bandwidth used for density estimation and is of order $o(N^{-2/5s})$ when s is the order of the kernel function employed[12], $\hat{E}_{D_g}$ and $\hat{V}_{D_g}$ are consistent estimates of

$$\begin{aligned} E_{D_g} &= \int\limits_u K^2(u)du \cdot \int\limits_\varepsilon [g(\varepsilon)]^2 \, d\varepsilon \\ V_{D_g} &= \int\limits_v \left[\int\limits_u K(u) \cdot K(u+v)du \right]^2 dv \cdot \int\limits_\varepsilon [g(\varepsilon)]^4 \, d\varepsilon, \end{aligned}$$

[12] A kernel function $K(u)$ is said to be of order s if its first $s-1$ moments are zero, while the s-th moment is finite and unequal to zero. The Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{u^2}{2}\right)$ is of order $s = 2$. In our empirical application, we used the bandwidth selector $h = 1.06 \cdot \hat{\rho}_{\hat{\varepsilon}} \cdot (\ln(N))^{-1} \cdot N^{-0.2}$, where N is the sample size and $\hat{\rho}_{\hat{\varepsilon}}$ is an estimate of the standard deviation of the estimated residuals $\hat{\varepsilon}$, as suggested to us by J. Grammig in personal communication.

and $K(\cdot)$ is the chosen Kernel function. The test is conducted as a one sided test so that large, positive values of FG lead to rejection of H_0.

As a second specification test we apply a method advanced by Diebold, Gunther, and Tay (1997) to test the forecast performance of the MSACD model. Denote by $\{f_n(x_n \mid \mathcal{F}_{n-1};\hat{\theta})\}_{n=1}^{N}$ the sequence of one-step-ahead density forecasts evaluated using parameter estimates $\hat{\theta}$ from some parametric model and by $\{f_n(x_n \mid \mathcal{F}_{n-1};\theta)\}_{n=1}^{N}$ the sequence of densities corresponding to the true, but unobservable data generating process of x_n. As shown by Rosenblatt (1952), under the null hypothesis

$$H_0: \quad \{f_n(x_n \mid \mathcal{F}_{n-1};\hat{\theta})\}_{n=1}^{N} = \{f_n(x_n \mid \mathcal{F}_{n-1};\theta)\}_{n=1}^{N} \tag{13}$$

the sequence of conditional empirical distribution functions defined by

$$\hat{\zeta}_n = \int\limits_{-\infty}^{x_n} f_n(u \mid \mathcal{F}_{n-1};\hat{\theta})\, du \tag{14}$$

is uniform i.i.d. on the unit interval. The recommendation of Diebold, Gunther, and Tay (1997) is to supplement statistical tests for i.i.d. uniformity by graphical tools. Departures from uniformity can easily be detected using a histogram plot based on the sequence of $\hat{\zeta}_n$ while the autocorrelogram for $\hat{\zeta}_n$ can be used in order to assess the maintenance of the i.i.d. property. By exploiting the statistical properties of the histogram under the null hypothesis of i.i.d. uniformity a straightforward goodness of fit test can be conducted. The ratio test RT_ζ

$$RT_\zeta \quad = \quad -2\ln\left(\prod_{k=1}^{K}\frac{\varsigma_k^{N_k}}{\hat{\varsigma}_k^{N_k}}\right) \;\sim\; \chi^2_{k-1}, \tag{15}$$

with N_k as the number of observations $\hat{\zeta}_n$ falling into the k^{th} bin, confronts the observed relative frequency $\hat{\varsigma}_k = \frac{N_k}{N}$ with the theoretical frequency ς_k. In order to test that all sample autocorrelations are simultaneously equal to zero one can apply the Ljung Box test statistic.

4 Empirical application

4.1 The data set

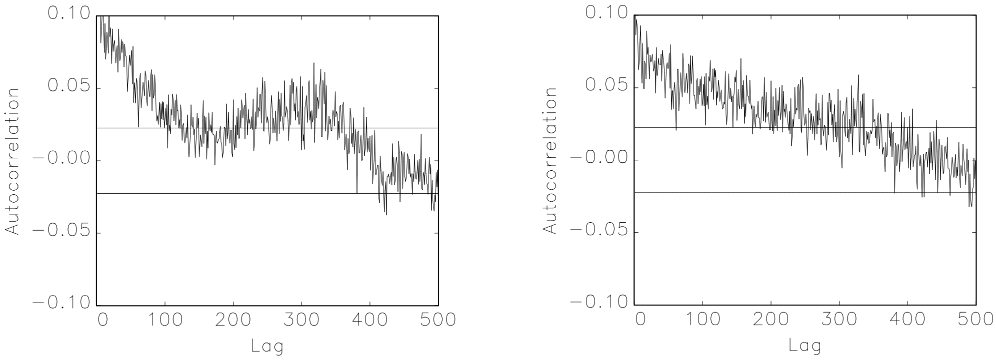
The data used in our empirical application consists of transactions of the common stock of Boeing recorded on the New York stock exchange (NYSE) from the trades and quotes database (TAQ) provided by the NYSE Inc. The sampling period spans 22 trading days from thursday August 1^{st} until friday August 30^{st}, 1996. We used all trades observed during the regular trading day (9:30 - 16:00). The trading times have been recorded with a precision measured in seconds. Observations occurring within the same second have been aggregated to one trade by summing the corresponding volumes and computing a volume weighted average of their prices. In the final data set we removed censored durations.[13]

Seasonality effects, such as the opening and closing of exchanges and lunch breaks, are meaningful reasons for strong dependence in transaction duration data. The deseasonalization of the raw durations and then the estimation of the model on the adjusted durations is a usual treatment of the seasonality in the duration process. A deseasonalized duration series x_n has been obtained by dividing the raw duratios $\tilde{x}_n$ through an appropriate estimate of the time of day function according to Eubank and Speckman (1990).[14] In contrast to this two step way, Veredas, Rodriguez-Poo, and Espasa (2002) propose an alternative approach for estimating jointly the duration dynamic and the intradaily seasonality.

Descriptive information about sample moments and Ljung Box statistics of the original and the seasonally adjusted duration data are reported in Table 1. As expected, the adjusted duration series has a mean of approximately one. Both time series exhibit overdispersion relative to the exponential distribution which has standard error equal to mean. Another characteristic of the data is the presence of strong, positive autocor-

[13]Durations from the last trade of the day until the close and durations from the open until the first trade of the day are declared to be censored.

[14]A Fourier series expansion accommodated by polynomials in the regressor variables is used to approximate the unknown time of day function.



Raw durations $\tilde{x}_n$ Adjusted durations x_n

Figure 1: Autocorrelation function for durations

relation in the trade durations. Even after seasonal adjustment, the Ljung Box tests for no autocorrelation up to 50 lags are rejected at the 5% significance level. Having regard to Figure 1 this means that there are periods with rare incidence of transactions implying successive long durations and periods with an amassment of transactions with very short time intervals between transactions. Therefore, an autoregressive approach appears to be appropriate as a model for the durations. In order to assess the out-of-sample forecast quality of the MSACD model, we divided our initial data set consisting of deseasonalized durations into two subperiods. The column titled "In-sample" contains the descriptive statistics for the first two thirds of the total sample which are used to estimate parameters. The rest of the data set is used to compute out-of-sample forecasts based on the estimated parameters. Durations in both subsamples appear to have similar characteristics, except for the occurrence of very large durations which tend to appear more concentrated in the first subsample.

4.2 The flexible scope of the MSACD model

For a given number of regimes equal to J, the parameter set $\theta_\psi^{(j)}$ determines the recursive linear or nonlinear regime specific mean function while $\theta_\varepsilon^{(j)}$ contains parameters which constitute the distribution of the residuals and the transition probabilities are gathered in θ_P. In the MSACD framework different restrictions on the parameter

Table 1: Descriptive Statistics for trade durations

Statistic	$\tilde{x}_n$	x_n	Subsamples	
			In-sample	Out-sample
Mean	67.7012	0.9993	0.9826	1.0326
Standard deviation	89.7997	1.2682	1.2625	1.2789
Overdispersion	119.1114	1.6095	1.6221	1.5839
Minimum	1.0000	0.0103	0.0127	0.0103
1st Quartile	13.0000	0.2045	0.1927	0.2320
Median	36.0000	0.5516	0.5433	0.5827
3rd Quartile	85.0000	1.2836	1.2526	1.3448
Maximum	1021.0000	15.5619	15.5619	10.6631
Interquartile range	72.0000	1.0791	1.0599	1.1127
N	7526	752	5016	2510
Ljung Box[a]	2747.1840	1887.3830	1704.5290	333.4175

[a] The Ljung Box statistic is based on 50 lags. For a significance level of 5% the tabulated critical value is 67.1671.

vector $\theta = (\theta_\psi^{(1)}, \ldots, \theta_\psi^{(J)}, \theta_\varepsilon^{(1)}, \ldots, \theta_\varepsilon^{(J)}, \theta_P)$ can be imposed.

First, the transition matrix P can be restricted in a way that the probability for regime j is independent of the state prevailed before, formally verbalized by $\pi_j \equiv p_{j1} = \ldots = p_{jJ}$. This implies a static mixture duration model which will be characterized by the restriction R_P. On the other hand, if we assume for each regime a particular duration process emerged from the same scope of a chosen ACD dynamic, different specifications also result from restrictions on the remaining distributional parameters and parameters describing the mean function. We refer to the specification R_ε when the regime specific distributional parameters are restricted to be equal by pairs, i.e. $\theta_\varepsilon^{(1)} = \ldots = \theta_\varepsilon^{(J)}$. When in addition the restriction $\theta_\psi^{(1)} = \ldots = \theta_\psi^{(J)}$ is imposed the resulting model has no different regime specific dynamics and the transition matrix stays unidentified. In such a case an ordinary ACD model can be used for data description. In the R_ψ specification where $\theta_\psi^{(1)} = \ldots = \theta_\psi^{(J)}$ is valid the mean function is independent of the regime in which the process resides.

In principle, the regime specific mean function can be parameterized in two ways. The simple variant S_ψ is characterized by the feature that lags of the regime specific conditional mean appear in the forecast function, yielding e.g. the following functional specification $\psi_n^{(j)} = \Upsilon\left(\psi_{n-1}^{(j)}, \psi_{n-2}^{(j)}, \ldots, \psi_{n-p}^{(j)}\right)$ while the more complex variant $\neg S_\psi$

includes lags of the regime independent conditional mean obtained by aggregation of regime specific means in the forecast function so that $\psi_n^{(j)} = \Upsilon(\psi_{n-1}, \psi_{n-2}, \ldots, \psi_{n-p})$. When the restriction R_ψ is imposed the specifications S_ψ and $\neg S_\psi$ are congruent. In the complex case $\neg S_\psi$ the pairwise identical regime specific mean functions implied through R_ψ are aggregated to a regime unspecific mean function ψ_n corresponding to $\psi_n^{(j)}$. In this case the time-consuming aggregation procedure is not necessary so that the choice of the simple variant S_ψ provides a great advantage in estimation.

The different restrictions in combination define a large repertory of MSACD models. For $J \geq 2$ each realization of the vector $(r_\psi, r_\varepsilon, r_P, s_\psi)$, with binary elements $r_z = 1$ if restriction R_z is imposed and $s_\psi = 1$ in the case of a simple mean specification S_ψ, determines one of all together 10 possible variants of interest.[15] When in addition the regime specific mean function $\psi_n^{(j)}$ is restricted to be independent from lagged means than both the simple variant S_ψ and the aggregation variant $\neg S_\psi$ effectuate the same estimation result. In this case the number of models shortens.

4.3 Estimation results

Beside the estimation of ordinary ACD models we estimated the MSACD model in its different variants for feasible number of regimes ($J = 2$ and $J = 3$). For each regime $j \leq J$ we assume a duration process emerged from an ordinary logarithmic ACD process based on the Burr distribution. More concrete, the regime specific mean function is of the form $\ln \psi_n^{(j)} = \omega^{(j)} + \sum_{k=1}^{p} \beta_k^{(j)} \ln(\psi_{n-1}^{(j)}) + \sum_{k=1}^{q} \alpha^{(j)} \ln(x_{n-1})$ in the simple case while it is $\ln \psi_n^{(j)} = \omega^{(j)} + \sum_{k=1}^{p} \beta^{(j)} \ln(\psi_{n-1}) + \sum_{k=1}^{q} \alpha^{(j)} \ln(x_{n-1})$ in the opposite case. Both, the lag orders p and q in the recursive mean functions are chosen to be zero or one. The regime specific distribution of the duration $x_n \mid s_n = j$ is chosen to be from the Burr family of distributions with distributional parameters $\kappa^{(j)}$, ${\sigma^{(j)}}^2$, and $\xi_n^{(j)}$ depending itself on $(\psi_n^{(j)}, \kappa^{(j)}, {\sigma^{(j)}}^2)$ so that $E[x_n \mid s_n = j, \mathcal{F}_{n-1}] = \psi_n^{(j)}$. The in-sample results of

[15] Effective, there are $2^4 = 16$ models. Because of congruence between $(1, 0, \cdot, 1)$ and $(1, 0, \cdot, 0)$ two specifications can be left out from estimation. Models with restrictions characterized by $(1, 1, \cdot, \cdot)$ do not account for regime specific dynamics so that in addition four models are superfluous.

the specification tests, values of the log-likelihood function and different information criteria for all models we estimated are gathered in Table 2.

Table 2: In-sample specification tests for MSACD models

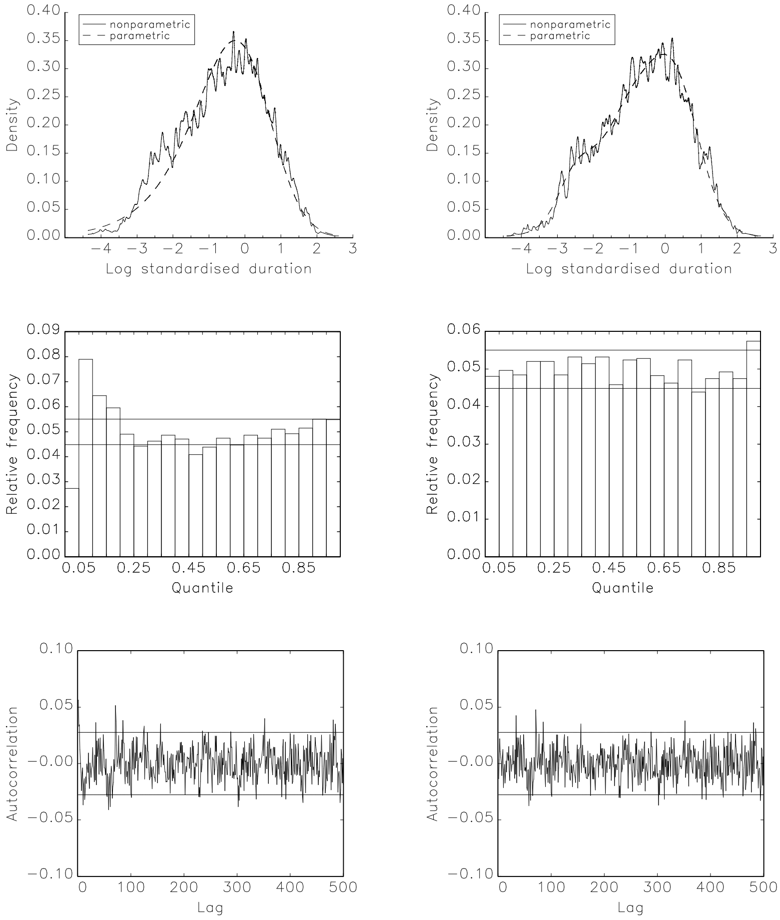
Lag orders	Restriction $(r_\psi, r_\varepsilon, r_P, s_\psi)$	$\ln \mathcal{L}_I$	AIC	BIC	$P(RT_\zeta)$	$P(LB_\zeta)$	$P(FG_\varepsilon)$	$P(LB_\varepsilon)$
				1 Regime Model				
$p=0,\ q=1$	$(1,1,\cdot,\cdot)$	-4698.34	9404.692	9430.774	0.0000	0.0000	0.0000	0.0000
$p=1,\ q=0$	$(1,1,\cdot,\cdot)$	-4764.74	9537.484	9563.566	0.0000	0.0000	0.0000	0.0000
$p=1,\ q=1$	$(1,1,\cdot,\cdot)$	-4556.32	9122.642	9155.244	0.0000	0.0013	0.0000	0.1417
				2 Regime Model				
$p=0,\ q=1$	$(0,1,1,\cdot)$	-4585.55	9185.101	9230.743	0.0000	0.0000	0.0000	0.0000
	$(1,0,1,\cdot)$	-4560.47	9134.958	9180.601	0.0291	0.0000	0.0308	0.0000
	$(0,0,1,\cdot)$	-4539.76	9097.523	9156.206	0.0104	0.0000	0.0049	0.0000
	$(0,1,0,\cdot)$	-4566.20	9148.403	9200.566	0.0000	0.0000	0.0045	0.0000
	$(1,0,0,\cdot)$	-4543.29	9102.587	9154.750	0.0000	0.0000	0.0000	0.0000
	$(0,0,0,\cdot)$	-4525.53	9071.062	9136.266	0.1360	0.0000	0.0013	0.0000
$p=1,\ q=0$	$(0,1,1,1)$	-4605.89	9225.793	9271.436	0.0000	0.0000	0.0000	0.0000
	$(0,1,1,0)$	-4642.15	9298.313	9343.956	0.0000	0.0000	0.0000	0.0000
	$(1,0,1,\cdot)$	-4576.69	9167.389	9213.032	0.0002	0.0000	0.0000	0.0000
	$(0,0,1,1)$	-4555.09	9128.194	9186.878	0.0001	0.0000	0.0000	0.0000
	$(0,0,1,0)$	-4603.60	9225.219	9283.903	0.0111	0.0000	0.0002	0.0000
	$(0,1,0,1)$	-4560.86	9137.732	9189.895	0.0000	0.0000	0.0000	0.0000
	$(0,1,0,0)$	-4569.87	9155.742	9207.905	0.0000	0.0000	0.0000	0.0000
	$(1,0,0,\cdot)$	-4543.77	9103.542	9155.705	0.0000	0.0000	0.0000	0.0000
	$(0,0,0,1)$	-4505.23	9030.465	9095.668	0.0001	0.0000	0.0014	0.0000
	$(0,0,0,0)$	-4443.05	8906.101	8971.305	0.0158	0.0000	0.0005	0.0000
$p=1,\ q=1$	$(0,1,1,1)$	-4452.30	8922.617	8981.301	0.0000	0.0008	0.0000	0.0984
	$(0,1,1,0)$	-4450.89	8919.789	8978.473	0.0000	0.0001	0.0000	0.1013
	$(1,0,1,\cdot)$	-4435.42	8886.844	8939.007	0.0949	0.0010	0.1972	0.0227
	$(0,0,1,1)$	-4408.98	8839.973	8911.697	0.1062	0.0006	0.0217	0.0755
	$(0,0,1,0)$	-4409.11	8840.228	8911.952	0.0418	0.0002	0.0046	0.0994
	$(0,1,0,1)$	-4442.56	8905.137	8970.341	0.0000	0.0561	0.0000	0.3319
	$(0,1,0,0)$	-4449.91	8919.822	8985.026	0.0000	0.0002	0.0000	0.0388
	$(1,0,0,\cdot)$	-4425.25	8868.518	8927.201	0.0312	0.1126	0.0299	0.0042
	$(0,0,0,1)$	-4399.72	8823.458	8901.702	0.0768	0.0577	0.0077	0.1453
	$(0,0,0,0)$	-4409.09	8842.193	8920.437	0.0639	0.0002	0.0039	0.0762
				3 Regime Model				
$p=0,\ q=1$	$(0,1,1,\cdot)$	-4525.67	9071.344	9136.548	0.0131	0.0000	0.3624	0.0000
	$(1,0,1,\cdot)$	-4549.95	9119.914	9185.118	0.5132	0.0000	0.7523	0.0000
	$(0,0,1,\cdot)$	-4507.99	9043.996	9135.282	0.5810	0.0000	0.9115	0.0000
	$(0,1,0,\cdot)$	-4494.05	9016.112	9107.397	0.0048	0.0000	0.0066	0.0000
	$(1,0,0,\cdot)$	-4469.29	8966.580	9057.866	0.8053	0.0000	0.0093	0.0000
	$(0,0,0,\cdot)$	-4445.72	8927.450	9044.817	0.7413	0.0000	0.9910	0.0000
$p=1,\ q=0$	$(0,1,1,1)$	-4559.05	9138.106	9203.310	0.0095	0.0000	0.0030	0.0000
	$(0,1,1,0)$	-4594.68	9209.373	9274.577	0.1749	0.0000	0.0619	0.0000
	$(1,0,1,\cdot)$	-4576.69	9173.389	9238.593	0.0002	0.0000	0.0000	0.0000
	$(0,0,1,1)$	-4527.61	9083.223	9174.509	0.0394	0.0000	0.1033	0.0000
	$(0,0,1,0)$	-4579.33	9186.669	9277.955	0.5158	0.0000	0.3803	0.0000
	$(0,1,0,1)$	-4509.79	9047.581	9138.867	0.0000	0.0000	0.0017	0.0000
	$(0,1,0,0)$	-4454.68	8937.366	9028.651	0.0001	0.0002	0.0007	0.0000
	$(1,0,0,\cdot)$	-4482.68	8993.372	9084.657	0.5433	0.0000	0.0893	0.0000
	$(0,0,0,1)$	-4449.65	8935.305	9052.672	0.8896	0.0000	0.8658	0.0000
	$(0,0,0,0)$	-4419.18	8874.362	8991.729	0.1166	0.0001	0.1613	0.0000
$p=1,\ q=1$	$(0,1,1,1)$	-4399.87	8825.755	8910.520	0.1058	0.0010	0.7599	0.2557
	$(0,1,1,0)$	-4393.43	8812.868	8897.633	0.1670	0.0003	0.2769	0.0477
	$(1,0,1,\cdot)$	-4435.42	8892.844	8964.569	0.1057	0.0007	0.1973	0.0179
	$(0,0,1,1)$	-4367.22	8768.452	8879.299	0.4762	0.0775	0.8628	0.7762
	$(0,0,1,0)$	-4388.76	8811.522	8922.368	0.4939	0.0028	0.4727	0.1001
	$(0,1,0,1)$	-4387.36	8808.724	8919.571	0.0433	0.0100	0.0680	0.2393
	$(0,1,0,0)$	-4386.41	8806.836	8917.682	0.6634	0.0005	0.0385	0.0616
	$(1,0,0,\cdot)$	-4389.28	8808.576	8906.381	0.4317	0.3380	0.2783	0.0000
	$(0,0,0,1)$	-4358.77	8759.542	8896.470	0.3914	0.1832	0.6620	0.5311
	$(0,0,0,0)$	-4368.00	8778.017	8914.945	0.9164	0.0005	0.1382	0.0655

$\ln \mathcal{L}_I$ is the value of the incomplete log-likelihood function, AIC is the Akaike's information criterion, computed as as $-2 \cdot \ln \mathcal{L}_I + 2 \cdot k$, where k is the number of estimated parameters, BIC is the Bayesian information criterion, computed as $-2 \cdot \ln \mathcal{L}_I + \ln(N) \cdot k$, where k is the number of estimated parameters, $P(RT_\zeta)$ is the p-value of the ratio test for the i.i.d. uniformity of ζ, using an histogram estimator for its density based on 20 equal bins, $P(LB_\zeta)$ is the p-value corresponding to the Ljung-Box statistic for 50 lags of ζ, $P(FG_\varepsilon)$ is the p-value of the nonparametric Fernandes and Grammig test statistic, $P(LB_\varepsilon)$ is the p-value corresponding to the Ljung-Box statistic for 50 lags of ε. All LB-statistics have been compared to critical values from a χ^2 distribution with $50 - (p + q + k)$ degrees of freedom where k is the number of estimated transition probabilities.

For a given number of regimes and a mean function chosen to be of order $p = 1$

and $q = 1$ the specifications implied by the restriction $(0, 0, \cdot, 1)$, i. e. the most flexible models with different regime specific mean functions without aggregation and regime specific distributional parameters, perform most suitable in terms of Akaike's information criterion AIC and Bayesian information criterion BIC. Note, that this plausible fact holds true either in the static mixture or in the Markov switching environment. In the static mixture type of the MSACD model with $p = 0$ and $q = 1$ and $J \in \{2, 3\}$, it is always the variant $(0, 0, 1, \cdot)$ which is outstanding with regard to AIC and BIC. In analogy, pure MSACD models with $p = 0$ and $q = 1$ perform best when no equality restrictions on $\theta_{\psi}^{(j)}$ and $\theta_{\varepsilon}^{(j)}$ are assumed. By way of an exception, it is the variant with aggregation procedure which is preferred by AIC or BIC if no lagged observed durations appear in the mean function ($p = 1$, $q = 0$) for either two or three regimes considered in the pure Markov switching model. Definitive, the three regime Markov switching model with lag orders $p = 1$ and $q = 1$ in the simple mean function and no restrictions on parameters at all is stamped by the lowest value of the AIC. Because of the abundance of parameters in the transition matrix the BIC supports the three regime static mixture model with $p = 1$ and $q = 1$ in the simple mean function and flexibility in the remaining parameters. Strictly, the AIC and BIC do not support the ordinary ACD models which are nested with $J = 1$ as special cases in the MSACD framework.

Also, none of the specification tests we discussed supports the ordinary ACD models. Neither the nonparametric FG test for the residuals ε_n nor the RT test for the integral transforms ζ_n is passed at conventional significance levels as the p-values of the corresponding test statistics indicate. From the plots of the density estimates of the residuals, as well as from the histogram of the series of integral transforms combined in Figure 2 on the left side, we find exemplary that a one regime duration model has severe problems to describe very small durations appropriately. Especially, the adjusted trade durations smaller than one contribute to this enormous misspecification in the one regime specification.



1-regime specification 3-regime specification

Figure 2: Results of the specification tests for 1-regime versus 3-regime MSACD models with $p = 1$ and $q = 1$. First row: Estimates of the density of the log residuals and corresponding theoretical density of log residuals implied by the estimated in-sample model. Second row: Histogram plots of the cumulative forecast density and 90% confidence intervals. Third row: Autocorrelation function and 95% confidence band for integral transforms.

In the two regime case for lag orders $p = 1$ and $q = 1$ in the regime specific mean functions, models with distributional parameters restricted to be equal across regimes, i. e. $(0, 1, \cdot, \cdot)$-specifications, have p-values of RT_ζ and FG_ε equal to zero. This means that there are further on grave deficiencies to capture distributional features of the duration process while the enormous autocorrelation in the adjusted data data can be eliminated to a greater or lesser extent as the high p-values of the Ljung Box statistics for the residuals ε_n and integral transforms ζ_n indicate. In the opposite case when only parameters determining the mean functions are restricted to be equal across regimes, i. e. $(1, 0, \cdot, \cdot)$-specifications, the hypothesis of i.i.d. uniformity in the ζ-series cannot be rejected at a significance level not less than 3%. Similarly, the null hypothesis that the unknown true distribution of the residuals equals their parametric distribution implied by the model cannot be rejected at a significance level not less than 3% again. This means that the enormous distributional problems of the origin ACD models (one regime models) can be removed in a satisfactory way when a mixture distribution for the duration data is assumed and in general the hypothesis of no autocorrelation in the ζ- and ε-series cannot be rejected as well. But more frugal specifications with either $p = 0$ or $q = 0$ in the mean functions, characterized by the restrictions $(1, 0, \cdot, \cdot)$ and $(0, 1, \cdot, \cdot)$ respectively, come off very badly even in sense of no remaining autocorrelation in the duration process as it can be seen from the extremely low p-values of the Ljung Box statistics.

Within the scope of three regime models specifications with restricted distributional parameters, i.e variants of the form $(0, 1, \cdot, \cdot)$, do not necessarily cause a mouldy goodness of fit. It will change for the better when at least the distributional parameters $\kappa^{(j)}$ and ${\sigma^{(j)}}^2$ of the chosen Burr distribution are allowed to be fully unrestrained. Therefor, the p-values of RT_ζ and FG_ε will raise up to 91%. The hypothesis of no autocorrelation in the residuals and integral transforms will be statistical significant as soon as $p = 1$ and $q = 1$ is assumed. The right side of Figure 2 demonstrates the excellence of the MSACD model for a specification with no distributional limitations and different regime specific dynamics in the mean functions without aggregation.

An objection raised to the use of chi-square tests is that information is thrown away by the grouping. So, it is not astonishing when sometimes the RT test supports the estimated model while at the same time it will be rejected through the FG test. There are two extreme cases, i. e. the specification $(0, 0, 0, \cdot)$ for $J = 2$ and the specification $(1, 0, 0, \cdot)$ for $J = 3$, both with lag orders $p = 0$ and $q = 0$ in the mean functions, where at a significance level of 10% the RT test retains the hypothesis of correct in-sample specification, but the FG test rejects it at the 1% significance level.

Table 3: Out-sample specification tests for MSACD models

Lag orders	Restriction $(r_\psi, r_\varepsilon, r_P, s_\psi)$	MSE	MAE	$P(RT_\zeta)$	$P(LB_\zeta)$	$P(FG_\varepsilon)$	$P(LB_\varepsilon)$
				1 Regime			
$p=0,\ q=1$	$(1,1,\cdot,\cdot)$	1.6201	0.8536	0.0000	0.0000	0.0000	0.0000
$p=1,\ q=0$	$(1,1,\cdot,\cdot)$	1.6359	0.8576	0.0000	0.0000	0.0006	0.0000
$p=1,\ q=1$	$(1,1,\cdot,\cdot)$	1.5939	0.8501	0.0002	0.3905	0.2310	0.4682
				2 Regime			
$p=0,\ q=1$	$(0,1,1,\cdot)$	1.6193	0.8637	0.0000	0.0000	0.0000	0.0000
	$(1,0,1,\cdot)$	1.6548	0.9324	0.0000	0.0000	0.0000	0.0000
	$(0,0,1,\cdot)$	1.6190	0.8605	0.0000	0.0000	0.0000	0.0000
	$(0,1,0,\cdot)$	1.6178	0.8607	0.0000	0.0000	0.0133	0.0000
	$(1,0,0,\cdot)$	1.6619	0.9358	0.0549	0.0011	0.0613	0.0000
	$(0,0,0,\cdot)$	1.6160	0.8760	0.0010	0.0000	0.0000	0.0000
$p=1,\ q=0$	$(0,1,1,1)$	1.6886	0.8487	0.0000	0.0000	0.0000	0.0000
	$(0,1,1,0)$	1.6360	0.8697	0.0000	0.0000	0.0000	0.0000
	$(1,0,1,\cdot)$	1.6956	0.9110	0.0000	0.0000	0.0001	0.0000
	$(0,0,1,1)$	1.6797	0.8454	0.0000	0.0000	0.0001	0.0000
	$(0,0,1,0)$	1.6359	0.8742	0.0000	0.0000	0.0000	0.0000
	$(0,1,0,1)$	1.6737	0.8409	0.0000	0.0000	0.0029	0.0000
	$(0,1,0,0)$	1.5969	0.8479	0.0217	0.2124	0.2769	0.0146
	$(1,0,0,\cdot)$	1.6800	0.9496	0.0246	0.0009	0.0060	0.0000
	$(0,0,0,1)$	1.6675	0.8368	0.0000	0.0000	0.1496	0.0000
	$(0,0,0,0)$	1.6004	0.8641	0.0000	0.3580	0.0000	0.0007
$p=1,\ q=1$	$(0,1,1,1)$	1.5967	0.8567	0.0013	0.3083	0.0000	0.2984
	$(0,1,1,0)$	1.5960	0.8552	0.0001	0.2349	0.0000	0.3620
	$(1,0,1,\cdot)$	1.6360	0.9217	0.0020	0.3685	0.0000	0.1328
	$(0,0,1,1)$	1.5950	0.8566	0.0000	0.3051	0.0000	0.2983
	$(0,0,1,0)$	1.5953	0.8485	0.0001	0.2976	0.0000	0.3871
	$(0,1,0,1)$	1.5920	0.8549	0.0007	0.4674	0.0000	0.3544
	$(0,1,0,0)$	1.5966	0.8558	0.0000	0.2814	0.0000	0.2956
	$(1,0,0,\cdot)$	1.6319	0.9196	0.0052	0.6080	0.0018	0.0196
	$(0,0,0,1)$	1.5921	0.8582	0.0001	0.5073	0.0000	0.2706
	$(0,0,0,0)$	1.5954	0.8490	0.0001	0.3049	0.0000	0.3415
				3 Regime			
$p=0,\ q=1$	$(0,1,1,\cdot)$	1.6195	0.8496	0.0000	0.0000	0.0000	0.0000
	$(1,0,1,\cdot)$	1.7219	0.9923	0.0002	0.0000	0.0001	0.0000
	$(0,0,1,\cdot)$	1.6198	0.8493	0.0000	0.0000	0.0000	0.0000
	$(0,1,0,\cdot)$	1.5970	0.8579	0.0015	0.0245	0.0470	0.0001
	$(1,0,0,\cdot)$	1.6956	0.9744	0.0257	0.0692	0.0005	0.0000
	$(0,0,0,\cdot)$	1.5967	0.8435	0.0080	0.0562	0.0047	0.0010
$p=1,\ q=0$	$(0,1,1,1)$	1.6872	0.8418	0.0000	0.0000	0.0096	0.0000
	$(0,1,1,0)$	1.6371	0.8538	0.0000	0.0000	0.0000	0.0000
	$(1,0,1,\cdot)$	1.6956	0.9110	0.0000	0.0000	0.0001	0.0000
	$(0,0,1,1)$	1.6738	0.8162	0.0000	0.0000	0.0000	0.0000
	$(0,0,1,0)$	1.6362	0.8664	0.0007	0.0000	0.0013	0.0000
	$(0,1,0,1)$	1.5940	0.8556	0.0486	0.0863	0.0253	0.0000
	$(0,1,0,0)$	1.5972	0.8616	0.0007	0.1867	0.0000	0.0019
	$(1,0,0,\cdot)$	1.7711	1.0269	0.0763	0.0597	0.0002	0.0000
	$(0,0,0,1)$	1.5992	0.8462	0.0153	0.1304	0.1503	0.0005
	$(0,0,0,0)$	1.5817	0.8610	0.0001	0.3768	0.0000	0.2821
$p=1,\ q=1$	$(0,1,1,1)$	1.5939	0.8460	0.0009	0.2022	0.0369	0.2131
	$(0,1,1,0)$	1.5933	0.8490	0.0265	0.1814	0.1076	0.2497
	$(1,0,1,\cdot)$	1.6360	0.9217	0.0022	0.3303	0.0000	0.1124
	$(0,0,1,1)$	1.5885	0.8475	0.0057	0.2288	0.0147	0.1301
	$(0,0,1,0)$	1.6176	0.8996	0.0021	0.2575	0.0037	0.3638
	$(0,1,0,1)$	1.5909	0.8492	0.0557	0.2165	0.3275	0.0608
	$(0,1,0,0)$	1.5956	0.8439	0.0038	0.2103	0.1898	0.2962
	$(1,0,0,\cdot)$	1.6468	0.9357	0.0205	0.7075	0.0633	0.0001
	$(0,0,0,1)$	1.5876	0.8447	0.0290	0.3790	0.6146	0.0046
	$(0,0,0,0)$	1.5963	0.8397	0.0169	0.2381	0.4425	0.3249

$MSE = N^{-1}\sum(x_n - \hat{\psi}_n)^2$, $MAE = N^{-1}\sum|x_n - \hat{\psi}_n|$. $P(RT_\zeta)$ is the p-value of the ratio test for the i.i.d. uniformity of ζ, using an histogram estimator for its density based on 20 equal bins, $P(LB_\zeta)$ is the p-value corresponding to the Ljung-Box statistic for 50 lags of ζ, $P(FG_\varepsilon)$ is the p-value of the nonparametric Fernandes and Grammig test statistic, $P(LB_\varepsilon)$ is the p-value corresponding to the Ljung-Box statistic for 50 lags of ε. All LB-statistics have been compared to critical values from a χ^2 distribution with $50-(p+q+k)$ degrees of freedom where k is the number of estimated transition probabilities.

In order to examine the suitability of forecasts the in-sample estimators have been applied to the out-sample trade duration data. Table 3 contains the p-values of several test statistics as well as the values of the mean squared error (MSE) and mean absolute error (MAE) which are used to form an opinion about the out-sample fit.

Except for one case, it comes out that for a given combination (p, q) in all regime specific mean functions there are always both two and three regime specifications with mean errors (MSE and MAE) smaller than the corresponding value in the origin ACD model. Starting from an ordinary ACD model, where lagged observed and expected durations determine the (simple) mean function ($p = 1$, $q = 1$), a slight improvement of the two forecast errors can be achieved when a crossover to a multiple regime environment will be executed with restricted distributional parameters. Reciprocal, a marginal deterioration of forecast performance is noticeable when the mean functions do not consist of regime specific dynamics. As a consequence, the predominance of the MSACD model is due to the introduction of a regime variable.

Concerning the hypothesis of i.i.d. uniformity for integral transforms ζ_n the standard ACD models provide p-values equal to zero. This finding persists also for static two and three regime models with mean functions characterized either by lag orders of $p = 0$, $q = 1$ or $p = 1$, $q = 0$. But the p-values $P(RT_\zeta)$ will increase up to 7.6% in the case of fully flexible transition probabilities for the Markov chain. In principle, this statement based on the RT test can be carried forward to the distributional features of the residuals. Even in the out-sample case, the MSACD framework is able to produce high p-values up to 61% for the hypothesis that the true but unknown distribution of the residuals ε_n equals a distribution implied by the estimated model.

As it can be seen from the p-values of the Ljung Box statistic, even for a moderate ACD model with $p = 1$ and $q = 1$ the hypothesis of no autocorrelation in the integral transforms and residuals cannot be rejected at a significance level more than 10%. This means that the mean dynamic is modelled adequately. Patterns of dependencies still exist when in principle a parsimonious specification of the mean function is chosen.

Recapitulating, the choice of the best model was based on the principle of parsimony and also on the ultimate goal to find a specification that yields a famous in-sample fit as well as a reasonable out-sample forecast performance for trade durations. With regard to the in-sample results, the three regime specification $(0, 0, 0, 1)$ with $p = 1$ and $q = 1$ preforms best in terms of the AIC, and the p-values of all specification tests are much greater than 10%. Furthermore this model also showed to be good in the out-sample forecast performance among all models that we considered as indicated by the extreme low values of the mean errors. And in general, the consideration of two or three regimes in the MSACD framework with adequate and sufficient regime specific dynamics yields reasonable results. This fact can be seen as a link to market microstructure theory. For example, following the sequential trade model of Easley, Kiefer, O'Hara, and Paperman (1996) the occurrence of different types of unobservable information events implies that trading evolves in different velocities and effectuates the price setting behavior of a market maker. It is based on the restrictive assumption that the arrival rate of sell orders under bad news equals the arrival rate of buy orders under good news. Furthermore it is assumed that buy and sell orders are equal when no news occur. From all, it follows that the data generating process of trade durations is a mixture of exponential distributions. This situation can be reproduced by a two regime MSACD model subject to extremely strict constrains in the parameter vector. Our application of the MSACD model can be regarded as a generalization of the sequential trade model which assumes that (i) the conditional density of the trade duration given the regime is not independent exponential but rather follows a logarithmic ACD model with marginal Burr density, (ii) the arrival rates are not restricted, and (iii) the information regime is not necessarily independent in time.

5 Conclusions

In this paper we proposed a new framework for modelling autocorrelated inter trade duration time series obtained from high frequency data sets from asset markets. The class of Markov switching models has been in use in econometrics for quite a while, but until now these models were based on marginal Gaussian processes. We showed that by analogy this framework may be used to estimate models based on non-Gaussian marginal distributions as well, and we described two alternative estimation techniques that may be employed in this context.

The MSACD model introduced in this paper was shown to be a successful tool for forecasting time series of intraday transaction durations. We showed that the MSACD model yields better in-sample fit and quite reasonable out-of-sample forecast performance compared to alternative ACD models. A further asset of the MSACD model is its interpretation in the context of recent market microstructure models.

Recently, the ACD-framework has been extended to the multivariate case as well (see Russell and Engle (1999) and Russell (1999)). A promising strategy for future research would be to combine the Markov switching approach with a multivariate extension of the ACD model. This would allow one to develop a more natural test of implications of many related microstructure models, as we might be able to explain the evolution of buyer and seller initiated trades as a bivariate duration process that depends on the unobservable stochastic information process.

References

Bauwens, L., F. Galli, and P. Giot (2001): "The moments of Log-ACD models," Discussion paper, CORE, Université Catholique de Louvain.

Bauwens, L., and P. Giot (2000): "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Thress NYSE Stocks," *Annales d'Économie et de Statistique*, 60, 118–149.

Bauwens, L., P. Giot, and D. Grammig, J. Veredas (2000): "A comparison of financial duration models via density forecasting," Discussion paper, Université Catholique de Louvain and University of Frankfurt, Discussion Paper.

Bauwens, L., and D. Veredas (1999): "The stochastic conditional duration model: A latent variable model for the analysis of financial durations," Université catholique de Louvain.

Böhning, D., E. Dietz, R. Schaub, Schlattman, and B. G. Lindsay (1994): "The distribution of the likelihood ratio of mixtures of densities from the one-parameter exponential family," *Annals of the Institute of Statistical Mathematics*, 46(2).

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 31(3), 307–327.

Box, G. E. P., and D. R. Cox (1964): "An analysis of transformations," *Journal of the Royal Statistical Society B (Methodological)*, 26, 211–243.

Cai, J. (1994): "A Markov model of switching-regime ARCH," *Journal of Business and Economic Statistics*, 12(3), 309–316.

Chow, G. C. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28(3), 591–605.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, 39, 1–38, Series B.

Dewachter, H. (2001): "Can Markov Switching models replicate chartist profits in the foreign exchange market ?," *Journal of International Money and Finance*, 20(1), 25–41.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1997): "Evaluating density forecasts," Discussion paper, University of Pennsylvania, Discussion Paper.

Easley, D., N. Kiefer, M. O'Hara, and J. P. Paperman (1996): "Liquidity, information and infrequently traded stocks," *Journal of Finance*, 51(4), 1405–1436.

Engel, C., and J. D. Hamilton (1990): "Long swings in the dollar: Are they in the data and do markets know it ?," *American Economic Review*, 80(4), 689–713.

Engle, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U. K. Inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and J. R. Russell (1998): "Autoregressive conditional duration: A new model for irregulary spaced transaction data," *Econometrica*, 66(5), 1127–1162.

Eubank, R. L., and P. Speckman (1990): "Curve Fitting by Polynomial-trigometric regression," *Biometrica*, 77(1), 1–9.

Fernandes, M., and J. Grammig (2000): "Non-parametric specification tests for conditional duration models," European University Institute and University of Frankfurt.

——— (2001): "A Family of Autoregressive Conditional Duration Models," Discussion paper, CORE.

Filardo, A. J. (1994): "Business-cycle phases and their transitional dynamics," *Journal of Business and Economic Statistics*, 12(3), 299–308.

Fong, W. M., and K. H. See (2001): "Modelling the conditional volatility of commodity index futures as a regime switching process," *Journal of Applied Econometrics*, 16, 133–163.

Francq, C., and J.-M. Zakoian (2001): "Stationarity of multivariate Markov-switching ARMA models," *Journal of Econometrics*, 102, 339–364.

Ghysels, E., C. Gouriéroux, and J. Jasiak (1997): "Stochastic volatility duration models," Working Paper 9746, CRESET, Paris.

Goldfeld, S., and R. Quandt (1965): "Some tests for homoscedasticity," *Journal of the American Statistical Association*, 60(310), 539–547.

——— (1973): "A markov model for switching regressions," *Journal of Econometrics*, 1(1), 3–16.

Grammig, J., and K.-O. Maurer (2000): "Non-monotonic hazard functions and the autoregressive conditional duration model," *The Econometrics Journal*, 3, 16–38.

Gray, S. (1996): "Modelling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27–62.

Hamilton, J., and R. Susmel (1994): "Autoregressive conditional heteroskedasticity and changes in regime," *Journal of Econometrics*, 64, 307–333.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57(2), 357–384.

——— (1990): "Analysis of time series subject to changes in regime," *Journal of Econometrics*, 45, 39–70.

——— (1994): *Time Series Analysis*. Princeton University Press.

Kim, C.-J. (1994): "Dynamic linear models with Markov-switching," *Journal of Econometrics*, 60(1), 1–22.

Lancaster, T. (1990): *The econometric analysis of transition data*. Cambridge University Press.

Lunde, A. (1999): "A Generalized Gamma Autoregressive Conditional Duration Model," Department of Economics, Politics and Public Administration, Aalbourg University, Denmark.

NELSON, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59(2), 347–370.

ROSENBLATT, M. (1952): "Remarks on a multivariate transformation," *Annals of Mathematical Statistics*, 23(3), 470–472.

RUSSELL, J. R. (1999): "Econometric Modeling of Multivaraite Irregularly-Spaced High-Frequency Data," Discussion paper, Graduate School of Business, University of Chicago, Discussion Paper.

RUSSELL, J. R., AND R. F. ENGLE (1999): "Econometric analysis of discrete-valued irregularly-spaced financial transaction data," Discussion paper, Graduate School of Business, University of Chicago, Discussion Paper.

VEREDAS, D., J. RODRIGUEZ-POO, AND A. ESPASA (2002): "On the (intradaily seasonality and dynamics of a financial point process: A semiparametric approach)," Discussion Paper 23, CORE, Université catholique de Louvain.

ZHANG, M. Y., J. R. RUSSELL, AND R. S. TSAY (2001): "A nonlinear Autoregressive Conditional Duration Model with applications to financial transaction data," *Journal of Econometrics*, 104(1), 179–207.