

Modeling the trading process on financial markets using the MSACD model

Reinhard Hujer ^{a,*}, Stefan Kokot ^b, Sandra Vuletić ^c

^a *University of Frankfurt/M., IZA Bonn*

^b *University of Frankfurt/M.*

^c *University of Frankfurt/M.*

Abstract

We propose a new framework for modeling time dependence in duration processes. The ACD approach introduced by Engle and Russell (1998) will be extended so that the conditional expectation of the durations depends on an unobservable stochastic process which is modeled via a Markov chain. The *Markov switching ACD model* (MSACD) is a flexible tool for description of financial duration processes. The introduction of a latent information 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 specific characteristics of inter trade durations while alternative ACD models fail.

Key words: Duration models, time series models, Markov switching models, financial transaction data, market microstructure.

JEL classification: C41, C22, C25, C51, G14.

* Corresponding author. Johann Wolfgang Goethe-University, Department of Economics and Business Administration, Institute of Statistics and Econometrics, Meritonstrasse 17, 60054 Frankfurt on the Main, Germany. Tel.: +49 69 798 28115; fax: +49 69 798 23673. *E-mail address:* hujer@wiwi.uni-frankfurt.de (R. Hujer).

Contents

1	Introduction	1
2	The ACD model	3
3	The Markov switching ACD model	4
3.1	The basic framework	4
3.2	Specification of the conditional mean	6
3.3	Stationarity conditions	8
3.4	Estimation of the MSACD model	9
3.5	Statistical inference	11
4	Empirical application	14
4.1	The data set	14
4.2	Specification of the MSACD Model	16
5	Testing implications of sequential trade models	20
6	Conclusions	23
A	Appendix	25
A.1	The distribution of the residuals	25
A.2	The Moving Average representation	26
	References	27
	Tables	30
	Figures	36

1 Introduction

The last twenty years saw an unprecedented upsurge in both theoretical and empirical work related to the analysis of market microstructure issues on financial markets.¹ Empirical studies are nowadays often based on *high frequency* data sets that contain detailed information about the timing of trades, prices, volume and other relevant characteristics for a wide range of financial securities. The availability of huge transaction data sets for academic research was accompanied by the introduction of new econometric methods which are tailor-made for the analysis of microstructure issues.

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 and econometric tools for the analysis of transition data and is therefore perfectly suited for the analysis of high frequency data sets which naturally arise as *irregularly spaced* data sets, i.e. the time between successive observations is not a deterministic constant but rather a random variable itself. ACD models have been almost exclusively used to analyze high frequency data stemming from stock and foreign exchange markets.

Following the seminal contribution of Engle and Russell (1998), a new branch in the econometric literature emerged quickly, that extended their original work in several directions. Despite the resulting variety of competing models, until now no satisfactory ACD variant in terms of forecast accuracy has been reported that could be used for the prediction of the trading process itself, see Bauwens et al. (2000). The main problem is the inability to forecast observations in the tails of their distribution appropriately.

Our intention is to introduce a reasonable statistical framework for time series of inter trade durations that can be used for forecasting purposes as well as for tests of the implications of market microstructure models. This will

¹ See Madhavan (2000) for a recent summary of this branch of literature.

be achieved by the introduction of a latent, discrete valued regime variable whose evolution in time is governed by a *Markov chain*. The inclusion of latent information structures in an ACD model can be justified in the light of several recent market microstructure models. The unobservable regime can be associated with the presence (or absence) of *private information* about an asset's value that is initially available exclusively to a subset of *informed traders* and only eventually disseminates through the mere process of trading to the broader public of all market participants. The *Markov switching ACD model* (MSACD) is closely related to the class of Markov switching autoregressive models introduced by Lindgren (1978) and Hamilton (1989). It provides a very flexible framework which nests many of the existing ACD models as special cases.

There are several extensions of the original ACD model that are related to our approach as well. The threshold ACD (TACD) model introduced by Zhang et al. (2001) allows switches between different regimes to be driven by past realizations of the dependent variable. Both the TACD and MSACD model belong to the class of discrete mixture models. ACD models based on a continuous mixture distribution are developed by Bauwens and Veredas (1999) and Ghysels et al. (2003).

This paper is structured as follows: Section 2 contains a brief review of the current state of art in ACD modeling. In Section 3 the MSACD model is introduced and compared to related work on duration models. Also, we discuss estimation procedures and specification tests for MSACD models. In an empirical application in Section 4 we present estimation results employing a transaction data set for the common share of Boeing traded on the New York Stock Exchange. The usefulness of the MSACD approach for tests of the implications of market microstructure models is demonstrated in Section 5 and finally, in Section 6 we summarize our main results and give a perspective on possible issues for future research.

2 The ACD model

The class of 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 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 or level of turnover has been traded.²

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

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

where \mathcal{F}_{n-1} may contain lagged dependent as well as lagged and contemporary exogenous variables, i.e. $\mathcal{F}_{n-1} = (x_1, \dots, x_{n-1}, y_1, \dots, y_n)$, and θ_ψ is the corresponding set of parameters. The ACD model is defined by some parameterization of this conditional mean and the decomposition

$$\varepsilon_n = \frac{x_n}{\psi_n}, \quad (2)$$

where the residual process ε_n is i.i.d. with density function $g(\varepsilon_n; \theta_\varepsilon)$ depending on a set of additional parameters θ_ε , support on the positive real line and an unconditional expectation equal to unity. The flexibility of the ACD model can be altered by modifying the distributional assumption of the residuals ε_n and/or the specification of the conditional mean function ψ_n . The distributional assumption of the residuals determines the density of the durations $f_n(x_n | \mathcal{F}_{n-1}; \theta)$ with $\theta = (\theta_\psi, \theta_\varepsilon)$ which will always belong to the same family of distributions as $g(\varepsilon_n; \theta_\varepsilon)$. A list of possible choices for $g(\varepsilon_n; \theta_\varepsilon)$ includes the exponential, the Weibull, the Burr, and the generalized gamma distribution. The exponential and Weibull ACD models originally introduced by Engle and Russell (1998) imply that the associated hazard rates are either constant or

² Naturally, the *price*, *volume* and *turnover* duration processes arise from the trade durations series by dropping intervening observations from the sample, thus yielding a 'thinned' or 'weighted' duration process.

monotonically increasing or decreasing. Added flexibility may be gained by specifying either a generalized gamma distribution (Lunde (1999)) or a Burr distribution (Grammig and Maurer (2000)).

In a standard ACD(p, q) model the parameterization of ψ_n is completely analogous to a GARCH model introduced by Engle (1982) and Bollerslev (1986)

$$\psi_n = \omega + \sum_{k=1}^p \beta_k \cdot \psi_{n-k} + \sum_{k=1}^q \alpha_k \cdot x_{n-k}, \quad (3)$$

and can be transformed into an ARMA ($\max(p, q), p$) representation from which expressions for the unconditional moments may be derived easily. In order to ensure non-negativity of ψ_n , the parameters ω , α_k , and β_k have to be non-negative as well. Bauwens and Giot (2002) circumvent this restriction by using the logarithmic LACD(p, q) specification

$$\ln(\psi_n) = \omega + \sum_{k=1}^p \beta_k \cdot \ln(\psi_{n-k}) + \sum_{k=1}^q \alpha_k \cdot \ln(x_{n-k}) \quad (4)$$

that closely resembles the EGARCH model of Nelson (1991). Analytical expressions for the unconditional moments of x_n in the LACD specification are given by Bauwens et al. (2003). In both specifications stationarity depends on the magnitudes of the parameters α_k , and β_k . Estimation of ACD models by maximum likelihood techniques is straightforward.

3 The Markov switching ACD model

3.1 The basic framework

The basic assumption of the MSACD model is that the conditional mean of the duration time series depends on an unobserved stochastic process s_n which represents the regime the process is in at time t_n . The interpretation of the regime variable usually varies with the specific aim of study. For example, in macroeconomic applications, 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 the existence of

different trading regimes may provide evidence on the presence of agents with private information about an assets's value. Thus, the dynamics of trading activity are different, depending on whether informed agents are active and on the nature of their information. The stochastic process s_n is a discrete valued random variable with support $\mathfrak{J} = \{j \mid 1 \leq j \leq J, J \in \mathbb{N}\}$.

The conditional mean of the durations x_n depends on the unobserved regime variable s_n in the following manner

$$\psi_n = \sum_{j=1}^J p(s_n = j \mid \mathcal{F}_{n-1}; \theta) \cdot \psi_n^{(j)}, \quad (5)$$

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 $\psi_n^{(j)} = E(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta)$ will have an autoregressive specification as in an ordinary ACD model. Decomposition (2) holds in the sense that the residuals ε_n have a known mixture distribution with $E(\varepsilon_n \mid \mathcal{F}_{n-1}) = 1$ and time-varying higher moments.³

The regime variable s_n switches between the states according to a Markov chain which is characterized by a $(J \times J)$ 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. The conditional density of the durations $f_n(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta)$ depends only on the current regime s_n , on the filtration \mathcal{F}_{n-1} , and on the parameter vector $\theta = (\theta_\psi^{(1)}, \dots, \theta_\psi^{(J)}, \theta_\varepsilon^{(1)}, \dots, \theta_\varepsilon^{(J)}, p_{11}, \dots, p_{JJ})'$. Any of the densities mentioned in Section 2 may be used. Since we cannot observe the realization of the current regime, the relevant density for statistical inference is the *marginal* density given by

$$f_n(x_n \mid \mathcal{F}_{n-1}; \theta) = \sum_{j=1}^J p(s_n = j \mid \mathcal{F}_{n-1}; \theta) \cdot f_n(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta). \quad (6)$$

Note that the MSACD model is a dynamic generalization of the static mixture hazard models introduced by Heckman and Singer (1982) with time varying mixture probabilities

³ See Appendix A.1.

$$\xi_{n+1|n}^{(j)} \equiv p(s_{n+1} = j \mid \mathcal{F}_n; \theta). \quad (7)$$

$\xi_{n+1|n}^{(j)}$ represents the ex-ante probability for being in regime j at time t_{n+1} , conditional on information available up to time t_n and can be evaluated using the two-step recursion ⁴

$$\xi_{n|n}^{(j)} = \frac{\xi_{n|n-1}^{(j)} \cdot f_n(x_n \mid s_n = j, \mathcal{F}_{n-1}; \theta)}{\sum_{k=1}^J \xi_{n|n-1}^{(k)} \cdot f_n(x_n \mid s_n = k, \mathcal{F}_{n-1}; \theta)} \quad (8)$$

$$\xi_{n+1|n}^{(j)} = \sum_{i=1}^J p_{ji} \cdot \xi_{n|n}^{(i)}. \quad (9)$$

Even though the transition probabilities p_{ji} are constant, the regime probabilities $\xi_{n|n}^{(j)}$ and $\xi_{n+1|n}^{(j)}$ are time-varying. A static mixture model in the spirit of Heckman and Singer (1982) may be regarded as a special case of the MSACD model based on a restricted transition matrix, where the elements of the j -th row are all equal, i.e. $\pi^{(j)} \equiv p_{j1} = \dots = p_{jJ}$. This implies time invariant forecasts of regime probabilities $\xi_{n+1|n}^{(j)} = \pi^{(j)}$ for all n but $\xi_{n|n}^{(j)}$ is still varying in time.

3.2 Specification of the conditional mean

Whenever the functional form of the conditional mean incorporates latent variables, such as lagged forecasts $\psi_{n-1}, \dots, \psi_{n-k}$ a problem of path dependence arises. Path dependence implies that the regime specific forecast depends on the entire sequence of regimes. This is so because the conditional mean ψ_n at time t_n depends on the conditional mean at time t_{n-1} which in turn depends upon the prevailing regime at time t_{n-1} . Therefore the distribution of x_n depends directly on s_n and also indirectly on the realizations for $(s_{n-1}, s_{n-2}, \dots, s_1)$. Since we cannot observe this sequence the likelihood function has to be constructed by integrating over all possible paths. An evaluation of all of the possible paths even for a moderate sample size is prohibitively expensive in terms of computational effort. One solution to avoid the problem of path dependence is to drop the impact of any latent variable such as lagged

⁴ See Hamilton (1994), pp. 692-694.

expected durations i.e. to consider only specifications with $p = 0$, as was done by Cai (1994) in the context of a model for the conditional variance. Alternatively, the problem can be avoided in a way that retains the important effect of persistence. There are in principle two ways in which lagged forecasts can appear in the conditional mean function $\psi_n^{(j)}$. In the *simple* model the current forecast $\psi_n^{(j)}$ is a function of lagged regime specific forecasts

$$\psi_n^{(j)} = \omega^{(j)} + \sum_{k=1}^p \beta_k^{(j)} \psi_{n-k}^{(j)} + \sum_{k=1}^q \alpha_k^{(j)} x_{n-k}. \quad (10)$$

Another possible specification is to make $\psi_n^{(j)}$ a function of past forecasts ψ_{n-k} that are regime independent as in the *complex* variant

$$\psi_n^{(j)} = \omega^{(j)} + \sum_{k=1}^p \beta_k^{(j)} \psi_{n-k} + \sum_{k=1}^q \alpha_k^{(j)} x_{n-k}. \quad (11)$$

It represents a 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

$$\psi_n = \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot \psi_n^{(j)}. \quad (12)$$

Both specifications given in (10) and (11) imply that the conditional mean depends only on the current regime, not on the entire past sequence of regimes. In both cases $\psi_n^{(j)}$ is the conditional mean at time t_n given that the process is in regime j . Note that specification (10) reaps an enormous benefit in comparison to the complex specification (11), since we may employ a variant of the Expectation-Maximization (EM) algorithm for estimation. Furthermore it implies that regime specific forecasts at t_n are being formed based on a comparison between past realized durations and regime specific forecasts. Specification (11) implies that forecasts are being formed based on a comparison between past realized durations and regime unspecific forecasts and involves more effort in estimation, since the EM-algorithm may not be employed.

3.3 Stationarity conditions

The specification of the mean function plays a crucial role for the stationarity conditions. For ARMA models subject to Markov switching it is standard to impose separate stationarity constraints for each regime, thus ensuring *local stationarity*. Because of the fact that both the ACD and the LACD models can be transformed into an ARMA representation, a set of local stationarity conditions for first order MSACD models will be⁵

$$|(\alpha^{(j)} + \beta^{(j)})| < 1 \quad \text{for } j = 1, \dots, J \quad (13)$$

Francq and Zakoïan (2001) have shown for Markov switching ARMA models that condition (13) can be relaxed. Their result can be used directly to derive stationarity conditions for the complex variant (11), which implies an autoregressive specification for x_n with time-varying coefficients

$$x_n = \omega_{n|n-1} - \beta_{n|n-1} \cdot \nu_{n-1} + (\alpha_{n|n-1} + \beta_{n|n-1}) \cdot x_{n-1} + \nu_n \quad (14)$$

where the innovation process $\nu_n \equiv x_n - \psi_n$ is characterized by the feature $E(\nu_n) = 0$, $\omega_{n|n-1} = \sum_{j=1}^J \xi_{n|n-1}^{(j)} \omega^{(j)}$, and $\alpha_{n|n-1}$ and $\beta_{n|n-1}$ are defined analogously to $\omega_{n|n-1}$. By successive recursion an infinite moving-average representation can be derived, leading to the relaxed condition for *global stationarity*

$$|\sum_{j=1}^J (\alpha^{(j)} + \beta^{(j)}) \cdot \pi^{(j)}| < 1 \quad (15)$$

with $0 \leq \pi^{(j)} \leq 1$ denoting the ergodic probability for regime j .⁶ Note that the existence of a regime specific unit root, i.e. $|\alpha^{(j)} + \beta^{(j)}| \geq 1$ for some $j \in \mathcal{J}$, does not necessarily violate (15). Local stationarity implies global stationarity but the opposite does not hold in general.

If the conditional mean function is specified according to the simple variant (10), the regime specific ARMA representation is given by

⁵ Generalizations to MSACD(p, q) models are straightforward. In order to simplify the notation in a first order MSACD model, we have dropped the indices k which determine the lag structure in the mean function.

⁶ See Appendix A.2.

$$x_n = \omega^{(j)} - \beta^{(j)}\nu_{n-1}^{(j)} + (\alpha^{(j)} + \beta^{(j)}) \cdot x_{n-1} + \nu_n^{(j)} \quad (16)$$

where the regime specific innovation $\nu_n^{(j)} = x_n - \psi_n^{(j)}$ has the property $E(\nu_n^{(j)} | s_n = j) = 0$. In this case global stationarity results from the local stationarity conditions (13).

3.4 Estimation of the MSACD model

In the case of regime switching models there are several ways in which maximum likelihood estimates of θ may be obtained. The standard approach maximizes directly the *incomplete* log-likelihood function $\mathcal{L}_I(\theta)$,

$$\ln \mathcal{L}_I(\theta) = \sum_{n=1}^N \ln [f_n(x_n | \mathcal{F}_{n-1}; \theta)] \quad (17)$$

numerically under the linear constraints $\sum_{k=1}^J p_{kj} = 1$ for all $j \in \mathcal{J}$ and additional restrictions for nonnegativity, stationarity and eventually for distributional parameters. The likelihood function for switching models may have more than one local maximum. It is therefore 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. Since standard maximization algorithms, such as the Newton-Raphson, often fail or produce nonsensical results, maximum likelihood estimates for Markov Switching models are often obtained by using the Expectation-Maximization (EM) algorithm introduced by Dempster et al. (1977) which is known for its numerical robustness.

The basis for the EM-algorithm is the hypothetical situation where we can observe the realization of the sequence of regime variables. Defining the random variables $z_n^{(j)} = 1$ if $s_n = j$ and zero otherwise, and $z_n^{(ji)} = z_n^{(j)} \cdot z_{n-1}^{(j)}$, the *complete* log-likelihood function $\ln \mathcal{L}_C(\theta)$ is given by⁷

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.

$$\begin{aligned} \ln \mathcal{L}_C(\theta) = & \sum_{n=1}^N \sum_{j=1}^J z_n^{(j)} \cdot \ln[f_n(x_n | s_n = j, \mathcal{F}_{n-1}; \theta)] \\ & + \sum_{n=2}^N \sum_{j=1}^J \sum_{i=1}^J z_n^{(ji)} \cdot \ln[p_{ji}]. \end{aligned} \quad (18)$$

The EM-algorithm proceeds by taking the expectation of (18) conditional on the observed data $\mathcal{X}_N = (x_1, \dots, x_N, y_1, \dots, y_N)$ and evaluates it using some arbitrary guess for the parameter vector θ_0 . The *expected complete* log-likelihood function $\ln \mathcal{L}_{EC}(\theta, \theta_0) \equiv E(\ln \mathcal{L}_C(\theta) | \mathcal{X}_N; \theta_0)$ is therefore given by replacing $z_n^{(j)}$ and $z_n^{(ji)}$ by appropriate probabilistic inferences $\xi_{n|N}^{(j)}$ and $\xi_{n|N}^{(ji)}$. These *smoothed inferences* may be evaluated employing a backward recursion starting with the filtered inferences $\xi_{N|N}^{(j)}$ obtained from (8) and progressing according to⁸

$$\xi_{n|N}^{(j)} = p(s_n = j | x_n, \mathcal{F}_N; \theta_0) = \xi_{n|n}^{(j)} \cdot \sum_{k=1}^J \frac{p_{kj} \cdot \xi_{n+1|n}^{(k)}}{\xi_{n+1|n}^{(k)}}. \quad (19)$$

and

$$\xi_{n|N}^{(ji)} = p(s_n = j, s_{n-1} = i | x_n, \mathcal{F}_N; \theta_0) = \xi_{n-1|n-1}^{(j)} \cdot \frac{p_{ji} \cdot \xi_{n|N}^{(j)}}{\xi_{n|n-1}^{(j)}}. \quad (20)$$

Evaluation of $\ln \mathcal{L}_{EC}(\theta, \theta_0)$ constitutes the first part of the EM-algorithm and is commonly referred to as the *E-step*. The associated *M-step* consists of maximizing $\ln \mathcal{L}_{EC}(\theta, \theta_0)$ with respect to the parameter vector θ , and can be conducted separately with respect to the parameters of the ACD model and the transition probabilities, if $\frac{\partial f_n(x_n | s_n = j, \mathcal{F}_{n-1}; \theta)}{\partial p_{mk}} = 0$ for all $j, m, k \in (1, \dots, J)$. The first order conditions lead to the following estimator for the transition probabilities

$$\hat{p}_{ji} = \frac{\sum_{n=2}^N \xi_{n|N}^{(ji)}}{\sum_{n=2}^N \xi_{n-1|N}^{(i)}}. \quad (21)$$

The remaining parameters may be obtained from the solution to

⁸ This algorithm has been proposed by Kim (1994). It is valid only when the regime variable 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} .

$$\sum_{n=1}^N \sum_{j=1}^J \xi_{n|N}^{(j)} \cdot \left(\frac{\partial \ln f_n(x_n | s_n = j, \mathcal{F}_{n-1}; \theta)}{\partial \theta} \right) \stackrel{!}{=} 0. \quad (22)$$

The parameters associated with the j -th regime, i.e. $\theta^{(j)} = (\theta_\psi^{(j)}, \theta_\varepsilon^{(j)})'$, may be estimated independently, if $\frac{\partial f_n(x_n | s_n = j, \mathcal{F}_{n-1}; \theta)}{\partial \theta^{(k)}} = 0$ for all $k \neq j$. Thus, by repeating the two steps of the EM-algorithm until the absolute change of the parameter vector is smaller than some prespecified convergence criterion estimates of the parameter vector are obtained. Hamilton (1990) shows that the final estimates $\hat{\theta}$ maximize both the expected complete log likelihood function as well as the incomplete log likelihood function.

3.5 Statistical inference

When conducting specification tests in Markov switching models, some care has to be exercised in order to avoid incorrect decisions as a result of the non-standard distributions of the test statistics involved. An example is testing whether a given data set may be described by a J -regime model or whether $(J - 1)$ regimes are sufficient. As shown by Böhning et al. (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 J 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, see Lindgren (1978). 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 is significantly different from zero may be compared to tabulated critical values of the t -distribution. Leroux (1992) shows that if the number of regimes is unknown a priori it can be determined consistently

by using information criteria, e.g. the Bayesian information criterion *BIC* proposed by Schwarz (1978).

Fernandes and Grammig (2000) have introduced a specification test for ordinary ACD models which is based on the discrepancy between the observed and the theoretical density function of the residuals and is, with minor refinements, applicable to the MSACD model as well. In ordinary ACD models the test statistic is easily derived 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

$$H_0 : \exists \theta \in \Theta \text{ such that } g(\varepsilon; \theta) \equiv \frac{1}{N} \cdot \sum_{n=1}^N g_n(\varepsilon | \mathcal{F}_{n-1}; \theta) = g(\varepsilon) \quad (23)$$

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

$$D_g = \frac{1}{N} \sum_{n=1}^N \left[g(\hat{\varepsilon}_n; \hat{\theta}) - \hat{g}(\hat{\varepsilon}_n) \right]^2. \quad (24)$$

Under the null hypothesis (23) the statistic FG_ε has asymptotically a standard normal distribution. FG_ε is given by

$$FG_\varepsilon = \frac{N \cdot h^{0.5} \cdot D_g - h^{-0.5} \cdot \hat{E}_{D_g}}{\sqrt{\hat{V}_{D_g}}}, \quad (25)$$

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 employed⁹, \hat{E}_{D_g} and \hat{V}_{D_g} are consistent estimates of

⁹ 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 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 $\hat{\varepsilon}$.

$$E_{D_g} = \int_u K^2(u) du \cdot \int_\varepsilon [g(\varepsilon)]^2 d\varepsilon \quad (26)$$

$$V_{D_g} = \int_v \left[\int_u K(u) \cdot K(u+v) du \right]^2 dv \cdot \int_\varepsilon [g(\varepsilon)]^4 d\varepsilon, \quad (27)$$

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

As a second specification test we apply a method advanced by Diebold et al. (1998) to test the forecast performance of general dynamic models, that has been used by Bauwens et al. (2000) to evaluate different types of ACD models. Denote by $\{f_n(x_n | \mathcal{F}_{n-1}; \hat{\theta})\}_{n=1}^N$ the sequence of one-step-ahead density forecasts evaluated using parameter estimates $\hat{\theta}$ from some parametric model and by $\{f_n(x_n | \mathcal{F}_{n-1}; \theta)\}_{n=1}^N$ 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

$$H_0 : \{f_n(x_n | \mathcal{F}_{n-1}; \hat{\theta})\}_{n=1}^N = \{f_n(x_n | \mathcal{F}_{n-1}; \theta)\}_{n=1}^N \quad (28)$$

the sequence of conditional empirical distribution functions (integral transforms) defined by

$$\hat{\zeta}_n = \int_{-\infty}^{x_n} f_n(u | \mathcal{F}_{n-1}; \hat{\theta}) du \quad (29)$$

is uniform i.i.d. on the unit interval. Therefore, any test for uniformity of the sequence of integral transforms $\hat{\zeta}_n$ can be used to assess the forecast performance of the model under consideration. Consider partitioning the support of ζ_n into K equally spaced bins and denote the number of observations falling into the k -th bin by N_k . The test statistic RT_ζ

$$RT_\zeta = -2 \ln \left[\prod_{k=1}^K \frac{\zeta_k^{N_k}}{\hat{\zeta}_k^{N_k}} \right] \quad (30)$$

compares the observed relative frequency $\hat{\zeta}_k = \frac{N_k}{N}$ to the theoretical frequency $\zeta_k = \frac{1}{K}$ and has a χ^2 distribution with $(K - 1)$ degrees of freedom under the null hypothesis. Additionally, the independence feature may be tested by computing the Ljung and Box (1978) test for the sequence of empirical

integral transforms $\hat{\zeta}_n$. The recommendation of Diebold et al. (1998) 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 independence property.

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 19 trading days from November 1 to November 27, 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 two kinds of censored durations: Durations from the last trade of the day until the close and from the open until the first trade of the day.

It is well known that the length of the durations varies in a deterministic manner during the trading day that resembles an inverted U-shaped pattern, i.e. intensity is very high after the open and before the close while it tends to be low during the middle of the day. Engle and Russell (1997) propose to decompose the duration series into a deterministic function of the time of day $\Phi(t_{n-1})$ and a stochastic component x_n , so that the raw durations are equal to $\tilde{x}_n = x_n \cdot \Phi(t_{n-1})$. In order to remove the deterministic component we apply the two step method proposed by Engle and Russell (1997) in which the time of day function is estimated separately from other model parameters.¹⁰

¹⁰ Simultaneous ML-estimation as in Engle and Russell (1998) and Veredas et al. (2002) is also feasible. Engle and Russell (1998) report that both procedures give similar results if sufficient data is available.

Dividing each raw duration \tilde{x}_n in the sample by an estimate of the time of day function $\Phi(t_{n-1})$, a sequence of deseasonalized durations x_n is obtained that is used in all subsequent analysis.¹¹

Descriptive information about sample moments and Ljung Box statistics of the raw and the seasonally adjusted duration data is reported in Table 1.

< insert Table 1 about here >

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 autocorrelation in the trade durations as can be seen in Figure 1.

< insert Figure 1 about here >

Even after seasonal adjustment, the Ljung-Box tests for no autocorrelation up to 50 lags are rejected at the 5% significance level, although the shape of the ACF changes slightly. 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 divide our initial data set consisting of 9092 observations into two subperiods. The column titled "In-sample" contains the descriptive statistics for the first 6060 observations (corresponding to two thirds of the total sample) which are employed to estimate parameters used for forecast evaluation. The rest of the data set is used to compute out-of-sample forecasts based on the estimated parameters. Descriptive statistics for the second subsample are contained in the column named "Out-sample". Durations in both subsamples appear to have similar

¹¹ Estimates of the time of day function were obtained by conducting a semi-nonparametric regression of the durations on the time of day according to Gallant (1981) and Eubank and Speckman (1990). Details on the seasonality adjustment step are available from the authors upon request.

characteristics, except for the occurrence of very large durations, which tend to appear more concentrated in the first subsample.

4.2 Specification of the MSACD Model

We estimate an ordinary ACD model and several MSACD model specifications with two, three and four regimes. We focus on the class of logarithmic MSACD models and distinguish between the two different specifications of the mean function introduced in Section 3.2. The simple variant (10), denoted by S in the following, may be estimated employing the EM-algorithm, while the complex variant (11), denoted by \bar{S} in the following, has to be estimated by maximization of the incomplete log-likelihood function. In both cases the lag orders p and q in the recursive mean functions are equal to one. Each regime specific distribution is chosen to be from the Burr family of distributions with time-invariant distributional parameters $\kappa^{(j)}$, $\sigma^{(j)}$, and a time-variant parameter $\xi_n^{(j)} = \psi_n^{(j)} \cdot \aleph^{(j)}$, with

$$\aleph^{(j)} = \frac{\sigma^{(j)} \left(1 + \frac{1}{\kappa^{(j)}}\right) \cdot \Gamma\left(\frac{1}{\sigma^{(j)}} + 1\right)}{\Gamma\left(1 + \frac{1}{\kappa^{(j)}}\right) \cdot \Gamma\left(\frac{1}{\sigma^{(j)}} - \frac{1}{\kappa^{(j)}}\right)} \quad (31)$$

so the regime specific density of x_n is given by

$$f_n(x_n | s_n = j, \mathcal{F}_{n-1}; \theta) = \frac{\xi_n^{(j)-\kappa^{(j)}} \cdot \kappa^{(j)} \cdot x_n^{\kappa^{(j)}-1}}{\left(1 + \sigma^{(j)} \cdot \xi_n^{(j)-\kappa^{(j)}} \cdot x_n^{\kappa^{(j)}}\right)^{\frac{1}{\sigma^{(j)}}+1}}. \quad (32)$$

The regime specific expectation is equal to $E[x_n | s_n = j, \mathcal{F}_{n-1}] = \psi_n^{(j)}$.

The in-sample results of the specification tests, values of the log-likelihood function and information criteria for all of the model specifications we estimated are presented in Table 2.

< insert Table 2 about here >

The *BIC* does not support the ordinary ACD model which is nested as a special case in the MSACD framework when $J = 1$. Also, none of the specification

tests that we performed supports the one regime model. However, the ordinary ACD model is able to capture the autocorrelation pattern of the trade durations adequately as indicated by the p-value of the Ljung Box statistic for ζ_n as well as for ε_n .

In order to examine the accuracy of forecasts, the in-sample estimates of θ have been used to compute one step forecasts for the out-sample 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 .

< insert Table 3 about here >

All findings from the in-sample discussion also hold for the out-sample forecasting performance of the one regime model. Additionally, the one regime model performs bad in terms of values of the MSE and MAE . There are always multiple regime specifications with smaller forecast errors than the one regime ACD model.

The results for regime switching models indicate a significant improvement on the performance of the ordinary ACD model. An extensive assessment of the forecast performance of MSACD models is provided by Hujer et al. (2003), who show that the degree of improvement depends crucially on the type of restrictions imposed on the parameter vector. They conclude that even static mixture ACD models with regime independent dynamics in the mean function and regime specific distributional parameters perform reasonably well in terms of forecast accuracy.

For J greater than one, first order MSACD are able to eliminate the distributional problem of ordinary ACD models and the autocorrelation pattern in the duration data will be considered adequately. The p-values of the RT_ζ and FG_ε test will rise to over 10%. The hypothesis of no autocorrelation in the residuals and i.i.d. integral transforms will be statistical significant at conventional significance levels. For a given number of regimes, it is always the simple variant S that performs better in terms of the BIC and achieves lower

forecast errors. With regard to in-sample results, the simple specification S performs generally better than the corresponding \bar{S} version in terms of results of the specification tests that we conducted.

The choice of our preferred model was based on the principle of parsimony and also on our ultimate goal to find a model specification that yields a good in-sample fit as well as reasonable out of sample forecast performance for trade durations. The BIC prefers the 2-regime specification, but the results of the FG_ε tests do not support the 2-regime specification at all. Therefore, we focused on the 3-regime simple specification S , since this was the one that passed through all in-sample specification tests we conducted, while at the same time it is more parsimonious than the 4-regime model which is also reflected in lower values of the BIC . Furthermore, this model also showed the best out of sample forecast performance among all models that we considered as indicated by the low values of forecast errors. Even though the result of the FG_ε test for the out-sample does not support the 3-regime model with simple mean specification S , we find that it offers a reasonable compromise between in-sample and out-sample performance.

For purposes of comparison Figure 2 contains plots of the density estimates for $\ln[\hat{\varepsilon}]$, as well as the histograms for the series of integral transforms ζ for the 1-regime and the preferred 3-regime specification.

< insert Figure 2 about here >

The plots for the in-sample clearly show that the MSACD model produces forecast residuals that match the implied theoretical density very well and tends to give accurate forecasts over the whole range of observed values of x . In contrast, the plots for the one regime model show that estimates of the residual density disagree sharply with the theoretical density, and that it tends to produce systematically biased forecasts of small x , the histogram for the first four quantiles is outside of the 95% confidence interval. Out-sample plots for the one regime model confirm this picture, while the density plots for

three regime MSACD model reveal that the theoretical and estimated density of the residuals still seem to match quite well, but the variance of the kernel density estimates has increased substantially. Furthermore, the out of sample histogram estimates appear to be more erratic and occasionally lie clearly outside of the confidence interval. Even so, there is no sign of a systematic pattern of over- or underestimation as in the case of the one regime model.

Table 4 contains the corresponding parameter estimates and standard errors for the preferred three regime specification.

< insert Table 4 about here >

The parameter estimates for $\beta^{(j)}$ and $\alpha^{(j)}$ differ only marginally across the three regimes. But the sign of the constant term $\omega^{(j)}$ varies across the regimes, with a positive value in regime 2 and negative values in regimes 1 and 3. Furthermore, the second regime provides us with the smallest estimates for distributional parameters in comparison to other regimes. This has a strong impact on the shapes of regime specific hazard functions, as shown in Figure 3.

< insert Figure 3 about here >

While in all three regimes the hazard rate tends to rise rather quickly after a transaction has been observed, the hazard function under the second regime clearly gives substantively more weight to spells with a length of more than two units of time.

This corresponds nicely to the fact that regime two has the longest expected stay $D^{(j)} = (1 - p_{jj})^{-1}$, as well as the highest ergodic probability $\pi^{(j)}$ among all three regimes.¹² Roughly 53.1% of all transactions were generated in this regime, and it takes approximately 2.5 transactions on average to leave regime 2. The average length of stay and ergodic probabilities for regime 1 with

¹² $\pi^{(j)}$ is a function of the elements of the transition matrix P and can be interpreted as a long run forecast of the regime probability $\xi_{N+r|N}^{(j)}$ for $r \rightarrow \infty$, see Kim and Nelson (1999).

$D^{(1)} = 1.7$ and $\pi^{(1)} = 0.32$ as well as regime 3 with $D^{(3)} = 1.2$ and $\pi^{(3)} = 0.15$ are substantively lower. The parameter estimates for the entire sample of 9092 observations reproduced in Table 5 differ from the "In-sample" estimates only marginally, thus reinforcing the impression, that the chosen MSACD specification provides a robust model for the data generating process of the trade durations during the sample period under consideration.

< insert Table 5 about here >

It will be used to conduct tests of the implications of a market microstructure model.

5 Testing implications of sequential trade models

In the framework of Easley et al. (1996), henceforth denoted as EKOP, the price setting behavior of market makers is explained by differences of traders information sets with respect to future price movements. Their setup is a sequential model of the trading process which is driven by the interaction of two types of traders, *informed* who observe a signal indicating either that the asset is either overpriced (bad news) or underpriced (good news) or that there is no information on the assets true value (no news) and *uninformed traders* who do not observe any signal.

The trading behavior of informed traders will depend on the type of the information signal. When a low signal indicates bad news, the profit maximizing investment strategy will be to sell the asset, so the aggregate sell arrival rate will be higher than on a no news day, while on a good news day there will be a higher occurrence rate of buys. On a trading day without a news event all transactions result from the arrival of buy and sell orders from uninformed traders. The arrival rate of both, buy and sell orders by uninformed traders, is equal and assumed to be determined by an i.i.d. Poisson process. The buy and sell order arrival rates for informed traders are identical and governed by an i.i.d. Poisson process, which is independent of the behavior of uninformed

traders.

Note that the EKOP model implies, that trading evolves in different velocities, depending on the type of the signal that has been observed by informed traders. It also implies that the data generating process of trade durations will be a mixture of two i.i.d. exponential distributions, with mixture probabilities determined by the probabilities of the information regimes. The information regime itself is a latent random variable. Thus, the MSACD model may be regarded as a generalization of the EKOP model, in which it is assumed that the information regime is not independent in time, but evolves according to a Markov chain during the trading day, the conditional densities of the trade durations given the regime are not independent exponentials but rather follow a (logarithmic) ACD model, with marginal Burr density. Furthermore if the arrival rates of either uninformed traders or informed traders (or both) are not restricted to be the same for buy and sell orders, the data generating process of the trade durations will be a three regime mixture model.

Another implication of the EKOP model, that we would expect to be consistent with our generalization, is that the occurrence of buyer and seller initiated transactions depends on the information regime. We therefore propose to test this implication of the EKOP-model by running an auxiliary regression of the type

$$\tilde{b}_n = \gamma + \phi \cdot \cos(h(t_n)) + \delta \cdot \sin(h(t_n)) + \sum_{j=1}^{J-1} \beta_j \cdot r_n^{(i,j)} + \sum_{p=1}^P \varphi_p \cdot b_{n-p}, \quad (33)$$

where $\tilde{b}_n = p(b_n = 1)$ is the probability, that the n -th observed trade is buyer initiated, $r_n^{(i,j)} = \ln(\xi_{n|N}^{(i)}) - \ln(\xi_{n|N}^{(j)})$, $\xi_{n|N}^{(j)}$ is the smoothed inference on the state of the regime variable s_n implied by the estimated MSACD model, b_n is an indicator variable, which is equal to one, if the n -th transaction was buyer initiated, and equal to zero, if it was seller initiated¹³ and the sine and

¹³We employ the 'quote test' proposed by Lee and Ready (1991) to determine the trade direction. This algorithm compares trade prices to the prevailing bid and ask prices. If trades occur before quotes are posted, the quote test compares the actual trade price to lagged trade prices, but if the trading day starts with a sequence of trades at the same price, it is not possible to classify them unambiguously. In our

cosine terms are included in order to control for deterministic time of day effects in the occurrence rates of buys and sells, with normalizing function $h(t_n) = 2\pi \cdot (t_n - t_{\min}) \cdot (t_{\max} - t_{\min})^{-1}$. t_{\min} (t_{\max}) is the time of day at which trading begins (ends) at the NYSE.

The inclusion of lagged b_n helps to account for possible strategic behavior of the informed traders, who may be reluctant to trade large quantities of the stock in a single trade, but rather prefer to split trades during the trading day. It is well known, that trades with large quantities have higher price effects than small trades, and thus, strategic order placement by informed traders might help them to hide their information as long as possible.¹⁴

Our specification stresses the magnitudes of the probability of being in regime i relative to the probability of being in regime j as the main determinant of the inclination to buy. Note that by comparing the magnitude and the sign of the two β -coefficients we are able to identify the nature of the information regime unambiguously. A positive coefficient of $r_n^{(1,2)}$ implies that the inclination to buy will increase, whenever $\xi_{n|N}^{(1)} > \xi_{n|N}^{(2)}$. If additionally the coefficient of the log ratio of regime 1 and 3 has a negative sign, then regime 1 is the no news regime, regime 2 is the bad news regime and regime 3 is the good news regime. Since the dependent variable is qualitative in nature, we estimate the parameter vector of the regression function employing the probit model. In order to find a reasonable specification for the regression function, we tried several different model specifications, see Table 6.

< insert Table 6 about here >

The three specifications differ only with respect to the inclusion of explanatory variables, with model 1 including only a constant and the log ratios $r_n^{(1,2)}$,

sample of transactions there were 25 trades in total that could not be classified, so the sample sizes for the regressions conducted in this section differ from those in the last section.

¹⁴ Another explanation for time dependence of the b_n sequence is herding behavior induced by strategic considerations of uninformed traders, who condition their own trades on the observed order flow.

model 2 additionally includes sine and cosine terms, and model 3 includes lags of b_n in the regression function.¹⁵

Model 3 provides the best fit to the data, when judged by the magnitudes of the R^2 goodness of fit measures and the value of the BIC . When lagged b_n are included, the sine and cosine terms tend to become insignificant, but note that the coefficients of the log ratios in all three models are significantly different from zero. Also, all three specifications imply, that the first regime is the good news regime (since, the coefficient of the log ratio of regime 2 and 1 is always negative, implying, that a higher probability of being in regime 1 than in regime 2 increases the probability of observing a buy), while regime 2 is the no news regime and regime 3 is associated with bad news.

Another quantity of interest is the probability of informed trading, that is implied by the parameter estimates of the EKOP model. The corresponding quantities for our generalized version of the EKOP model can be derived from the ergodic probabilities of the Markov chain. For our preferred 3-regime MSACD model these are equal to $\pi^{(1)} = 0.2873$, $\pi^{(2)} = 0.5445$, and $\pi^{(3)} = 0.1682$, see Table 5. Thus the probability of informed trading in the sample period is equal to $1 - \pi^{(3)} = 0.4555$, while the probability of being in the good news regime 1 is roughly two times that of the bad news regime 2. These results nicely conform to our economic intuition, that the bulk of transactions results from order placement by uninformed traders, and that the November of 1996 basically saw a bull market for the common share of Boeing, with prices rising by 6.25% during our sample period.

6 Conclusions

In this paper we proposed a new framework for modeling 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 econo-

¹⁵ We included all significant lags of b_n in this specification. We also estimated specifications with higher order lags, but non of the corresponding parameter estimates appeared to be significant.

metrics 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 (1998) 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.

A Appendix

A.1 The distribution of the residuals in the MSACD model

Starting with the marginal density of x_n given in (6), which is a mixture distribution with expectation $\psi_n = \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot \psi_n^{(j)}$, the density of the residuals $\varepsilon_n \equiv \frac{x_n}{\psi_n}$ is equal to

$$g_n(\varepsilon_n | \mathcal{F}_{n-1}; \theta) = \psi_n \cdot \sum_{j=1}^J p(s_n = j | \mathcal{F}_{n-1}; \theta) \cdot f_x(\varepsilon_n \cdot \psi_n | s_n = j, \mathcal{F}_{n-1}; \theta), \quad (\text{A.1})$$

where $f_x(\cdot)$ denotes the density function of the durations x_n . The mean of ε_n is given by

$$E[\varepsilon_n | \mathcal{F}_{n-1}] = E\left[\frac{x_n}{\psi_n} | \mathcal{F}_{n-1}\right] = \frac{\psi_n}{\psi_n} = 1,$$

and thus independent of n as in a standard ACD model. Recall, that for a mixture density of the form $f(y) = \sum_{j=1}^J p(s = j) \cdot f_j(y | s = j)$ the raw (uncentered) moments μ'_m are given by¹⁶

$$\mu'_m = E(y^m) = \sum_{j=1}^J p(s = j) \cdot E(y^m | s = j).$$

In order to derive an expression for the variance of ε_n , we first define $Var(x_n | s_n = j, \mathcal{F}_{n-1}) \equiv \varrho_n^{(j)}$. In general the regime specific variance $\varrho_n^{(j)}$ will depend on the conditional distribution assumed for x_n . The uncentered second moment of x_n is equal to $E(x_n^2 | s_n = j, \mathcal{F}_{n-1}) = \varrho_n^{(j)} + (\psi_n^{(j)})^2$ and so the regime independent second moment is $E(x_n^2 | \mathcal{F}_{n-1}) = \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot (\varrho_n^{(j)} + (\psi_n^{(j)})^2)$. Thus the regime independent variance of x_n is

$$\begin{aligned} Var(x_n | \mathcal{F}_{n-1}) &= E(x_n^2 | \mathcal{F}_{n-1}) - [E(x_n | \mathcal{F}_{n-1})]^2 \\ &= \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot \varrho_n^{(j)} + \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot (\psi_n^{(j)})^2 - (\psi_n)^2. \end{aligned}$$

The variance of ε_n is a function of the moments of x_n and is equal to

$$\begin{aligned} Var(\varepsilon_n | \mathcal{F}_{n-1}) &= \frac{1}{\psi_n^2} \cdot Var(x_n | \mathcal{F}_{n-1}) \\ &= \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot \frac{\varrho_n^{(j)}}{\psi_n^2} + \sum_{j=1}^J \xi_{n|n-1}^{(j)} \cdot \left(\frac{\psi_n^{(j)}}{\psi_n}\right)^2 - 1 \\ &= \frac{1}{\psi_n^2} \cdot E(x_n^2 | \mathcal{F}_{n-1}) - 1. \end{aligned} \quad (\text{A.2})$$

¹⁶ See Cameron and Trivedi (1998), p. 130.

Thus, in general the variance of ε_n will change over time (and higher moments of ε_n also). From the expression in the second line of (A.2) a sufficient condition for time invariance of $\text{Var}(\varepsilon_n | \mathcal{F}_{n-1})$ is satisfied, when all the regime specific conditional means are equal ($\psi_n = \psi_n^{(j)}$) and the regime probabilities are independent of time ($\xi_{n|n-1}^{(j)} = \pi^{(j)}$). Expressions for higher order moments can be derived in the same manner.

A.2 The Moving Average representation

Starting from the ARMA representation

$$x_n = \omega_{n|n-1} - \beta_{n|n-1}\nu_{n-1} + (\alpha_{n|n-1} + \beta_{n|n-1})x_{n-1} + \nu_n \quad (\text{A.3})$$

of a linear first order MSACD process with complex specification of the mean function, successive recursion leads to an infinite sum

$$\begin{aligned} x_n = & \omega_{n|n-1} \\ & + \sum_{m=1}^{\infty} \omega_{n-m|n-m-1} \prod_{k=1}^m (\alpha_{n-k+1|n-k} + \beta_{n-k+1|n-k}) \\ & + (\nu_n - \beta_{n|n-1} \cdot \nu_{n-1}) \\ & + \sum_{m=1}^{\infty} (\nu_{n-m} - \beta_{n-m|n-m-1} \cdot \nu_{n-m-1}) \\ & \cdot \prod_{k=1}^m (\alpha_{n-k+1|n-k} + \beta_{n-k+1|n-k}) \end{aligned} \quad (\text{A.4})$$

Taking expectations of (A.4) under consideration that $E(\nu_n) = 0$, we see that

$$\begin{aligned} E(x_n) = & E(\omega_{n|n-1}) + \sum_{m=1}^{\infty} E(\omega_{n-m|n-m-1}) \\ & \cdot \prod_{k=1}^m \left(\sum_{j=1}^J (\alpha^{(j)} + \beta^{(j)}) \cdot E(\xi_{n-k+1|n-k}^{(j)}) \right) \\ = & \left(\sum_{j=1}^J \omega^{(j)} \cdot \pi^{(j)} \right) \cdot \sum_{n=0}^{\infty} \left(\sum_{j=1}^J (\alpha^{(j)} + \beta^{(j)}) \cdot \pi^{(j)} \right)^m \\ = & \frac{\sum_{j=1}^J \omega^{(j)} \cdot \pi^{(j)}}{1 - \sum_{j=1}^J (\alpha^{(j)} + \beta^{(j)}) \cdot \pi^{(j)}} \end{aligned} \quad (\text{A.5})$$

if $|\sum_{j=1}^J (\alpha^{(j)} + \beta^{(j)}) \cdot \pi^{(j)}| < 1$ is satisfied.

References

- Bauwens, L., Galli, F., Giot, P., 2003. The moments of log-ACD models. Discussion Paper 11, CORE, Université Catholique de Louvain.
- Bauwens, L., Giot, P., 2002. The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et Statistique* 60, 117–149.
- Bauwens, L., Giot, P., Grammig, J. Veredas, D., 2000. A comparison of financial duration models via density forecasts. Discussion Paper 60, CORE, Université Catholique de Louvain and University of Frankfurt, forthcoming in: *International Journal of Forecasting*.
- Bauwens, L., Veredas, D., 1999. The stochastic conditional duration model: A latent variable model for the analysis of financial durations. Discussion Paper 9958, Université Catholique de Louvain.
- Böhning, D., Dietz, E., Schaub, R., Schlattman, Lindsay, B. G., 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.
- Cai, J., 1994. A markov model of switching-regime arch. *Journal of Business and Economic Statistics* 12 (3), 309–316.
- Cameron, A. C., Trivedi, P. K., 1998. *Regression analysis of count data*. Cambridge University Press, Cambridge.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39, 1–38, series B.
- Diebold, F. X., Gunther, T. A., Tay, A. S., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39 (4), 863–883.
- Easley, D., Kiefer, N., O'Hara, M., Paperman, J. P., 1996. Liquidity, information and infrequently traded stocks. *Journal of Finance* 51 (4), 1405–1436.
- Engle, R. F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. inflation. *Econometrica* 50, 987–1008.
- Engle, R. F., Russell, J. R., 1997. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal*

- of Empirical Finance 4 (2-3), 187–212.
- Engle, R. F., Russell, J. R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66 (5), 1127–1162.
- Eubank, R. L., Speckman, P., 1990. Curve fitting by polynomial-trigonometric regression. *Biometrika* 77 (1), 1–9.
- Fernandes, M., Grammig, J., 2000. Non-parametric specification tests for conditional duration models. *EUI Working Paper in Economics* 4, European Institute.
- Fong, W. M., See, K. H., 2001. Modelling the conditional volatility of commodity index futures as a regime switching process. *Journal of Applied Econometrics* 16, 133–163.
- Francq, C., Zakoïan, J.-M., 2001. Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics* 102, 339–364.
- Gallant, A. R., 1981. On the bias in flexible functional forms and an essentially unbiased form. *Journal of Econometrics* 20 (2), 285–323.
- Ghysels, E., Gouriéroux, C., Jasiak, J., 2003. Stochastic volatility duration models. *Journal of Econometrics* (Forthcoming).
- Grammig, J., Maurer, K.-O., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal* 3, 16–38.
- Gray, S., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.
- Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57 (2), 357–384.
- Hamilton, J. D., 1990. Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.
- Hamilton, J. D., 1994. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.
- Heckman, J. J., Singer, B., 1982. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52 (2), 271–320.
- Hujer, R., Kokot, S., Vuletić, S., 2003. Comparison of MSACD models, Johann Wolfgang Goethe University, Frankfurt am Main.
- Kim, C.-J., 1994. Dynamic linear models with markov-switching. *Journal of Econometrics* 60 (1), 1–22.
- Kim, C.-J., Nelson, C. R., 1999. State-space models with regime switching. *The*

- MIT Press, Cambridge, Massachusetts.
- Lee, C. M. C., Ready, M. J., 1991. Inferring trade direction from intraday data. *Journal of Finance* 46 (2), 733–746.
- Leroux, B. G., 1992. Consistent estimation of a mixing distribution. *The Annals of Statistics* 20 (3), 1350–1360.
- Lindgren, G., 1978. Markov regime switching models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics* 5 (2), 81–91.
- Ljung, G. M., Box, G. E. P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65 (2), 297–303.
- Lunde, A., 1999. A generalized gamma autoregressive conditional duration model, Department of Economics, Politics and Public Administration, Aalborg University, Denmark.
- Madhavan, A., 2000. Market microstructure: A survey. *Journal of Financial Markets* 3 (3), 205–258.
- 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 multivariate irregularly-spaced high-frequency data, Graduate School of Business, University of Chicago.
- Russell, J. R., Engle, R. F., 1998. Econometric analysis of discrete-valued irregularly-spaced financial transaction data using a new autoregressive conditional multinomial model. UCSD Economics Discussion Papers 10, University of California at San Diego, Department of Economics.
- Schwarz, G., 1978. Estimating the dimension of a model. *Annals of Statistics* 6 (2), 461–464.
- Veredas, D., Rodriguez-Poo, J., Espasa, A., 2002. On the intradaily seasonality and dynamics of a financial point process: A semiparametric approach. Discussion Paper 23, CORE, Université Catholique de Louvain.
- White, H., 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1), 1–25.
- Zhang, M. Y., Russell, J. R., Tsay, R. S., 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* 104 (1), 179–207.

Tables

Table 1
Descriptive Statistics for trade durations

Statistic	\tilde{x}_n	x_n	Subsamples	
			In-sample	Out-sample
Mean	48.3248	1.0007	1.0435	0.9151
Standard deviation	61.8416	1.1933	1.2471	1.0727
Minimum	1.0000	0.0141	0.0141	0.0141
1st Quartile	10.0000	0.2323	0.2355	0.2272
Median	27.0000	0.5875	0.6014	0.5589
3rd Quartile	61.0000	1.2980	1.3739	1.1659
Maximum	894.0000	16.1672	16.1672	8.3896
Interquartile range	51.0000	1.0657	1.1384	0.9388
N	9092	9092	6060	3032
Ljung Box ^a	3815.6633	1362.7593	1018.9819	249.8529

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

Table 2
In-sample specification tests

Mean Specification	$\ln \mathcal{L}_I$	BIC	$P(RT_\zeta)$	$P(LB_\zeta)$	$P(FG_\varepsilon)$	$P(LB_\varepsilon)$
1 Regime model						
	-6025.59	12094.74	0.0000	0.3335	0.0000	0.6315
2 Regime model						
S	-5864.46	11833.44	0.1145	0.4048	0.0015	0.4905
\bar{S}	-5883.20	11870.92	0.0125	0.2562	0.0020	0.5505
3 Regime model						
S	-5828.55	11840.00	0.2666	0.1609	0.0581	0.1822
\bar{S}	-5840.06	11863.03	0.8391	0.2103	0.4758	0.7746
4 Regime model						
S	-5804.04	11886.78	0.4558	0.1188	0.7985	0.5552
\bar{S}	-5834.64	11947.99	0.0556	0.0194	0.0349	0.1278

$\ln \mathcal{L}_I$ is the value of the incomplete log-likelihood function, 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.

Table 3
Out-sample specification tests

Mean Specification	MSE	MAE	$P(RT_\zeta)$	$P(LB_\zeta)$	$P(FG_\varepsilon)$	$P(LB_\varepsilon)$
1 Regime model						
	1.1268	0.7268	0.0000	0.3802	0.0053	0.4840
2 Regime model						
S	1.1255	0.7252	0.0228	0.2724	0.0011	0.3382
\bar{S}	1.1271	0.7266	0.0296	0.2762	0.0002	0.3258
3 Regime model						
S	1.1227	0.7205	0.0124	0.0930	0.0000	0.1979
\bar{S}	1.1236	0.7271	0.0008	0.1466	0.0000	0.1527
4 Regime model						
S	1.1247	0.7192	0.0105	0.0156	0.0416	0.0640
\bar{S}	1.1257	0.7302	0.0008	0.0217	0.0003	0.0127

$MSE = N^{-1} \sum (x_n - \hat{\psi}_n)^2$, $MAE = N^{-1} \sum |x_n - \hat{\psi}_n|$. $P(RT_\zeta)$ 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.

Table 4
In-Sample estimation results

Parameter	Regime $j = 1$		Regime $j = 2$		Regime $j = 3$	
	Estimate	Stderr	Estimate	Stderr	Estimate	Stderr
Mean function						
$\omega^{(j)}$	-0.0102	0.0107	0.0243	0.0161	-0.0181	0.0053
$\alpha^{(j)}$	0.0215	0.0072	0.0238	0.0159	0.0031	0.0018
$\beta^{(j)}$	0.9663	0.0154	0.9757	0.0176	0.9883	0.0029
Distribution						
$\kappa^{(j)}$	2.1565	0.1570	1.6385	0.0776	3.4156	0.3910
$\sigma^{(j)}$	0.8749	0.1366	0.3559	0.0516	1.8981	0.3673
Probability						
p_{1j}	0.4112	0.0701	0.2572	0.0386	0.3232	0.0817
p_{2j}	0.4206	0.0821	0.5932	0.0377	0.5416	0.0943
p_{3j}	0.1682	-	0.1496	-	0.1352	-
$D^{(j)}$	1.6984	-	2.4582	-	1.1563	-
$\pi^{(j)}$	0.3160	-	0.5307	-	0.1533	-

Standard errors have been computed based on numerical derivatives of the incomplete log likelihood function using the quasi-maximum likelihood (QML) estimates of the information matrix as suggested by White (1982).

Table 5
Total sample estimation results

Parameter	Regime $j = 1$		Regime $j = 2$		Regime $j = 3$	
	Estimate	Stderr	Estimate	Stderr	Estimate	Stderr
Mean function						
$\omega^{(j)}$	-0.0165	0.0166	0.0187	0.0043	-0.0399	0.0167
$\alpha^{(j)}$	0.0262	0.0125	0.0178	0.0046	0.0077	0.0043
$\beta^{(j)}$	0.9511	0.0296	0.9812	0.0046	0.9741	0.0117
Distribution						
$\kappa^{(j)}$	2.1955	0.1228	1.6577	0.0600	3.0937	0.2745
$\sigma^{(j)}$	0.8665	0.1286	0.4080	0.0444	1.6414	0.2246
Probability						
p_{1j}	0.3889	0.0581	0.2211	0.0341	0.3279	0.0660
p_{2j}	0.4290	0.0607	0.6065	0.0382	0.5410	0.0849
p_{3j}	0.1821	-	0.1724	-	0.1311	-
$D^{(j)}$	1.6340	-	2.5413	-	1.1509	-
$\pi^{(j)}$	0.2873	-	0.5445	-	0.1682	-

Standard errors have been computed based on numerical derivatives of the incomplete log likelihood function using the quasi-maximum likelihood (QML) estimates of the information matrix as suggested by White (1982).

Table 6
 Estimation results for probit models

Variable	Specification 1		Specification 2		Specification 3	
	Estimate	<i>t</i> -value	Estimate	<i>t</i> -value	Estimate	<i>t</i> -value
Constant	0.1551	7.8313	0.1562	7.7974	-0.7724	-24.4609
$r_n^{(2,1)}$	-0.0843	-4.4836	-0.0835	-4.4281	-0.0484	-2.3643
$r_n^{(2,3)}$	0.0367	3.3643	0.0363	3.3200	0.0262	2.2527
$\cos(h(t_n))$	-	-	-0.0322	-1.6852	-0.0117	-0.5685
$\sin(h(t_n))$	-	-	0.0380	1.9913	0.0193	0.9396
b_{n-1}	-	-	-	-	0.9185	29.5184
b_{n-2}	-	-	-	-	0.3297	9.9441
b_{n-3}	-	-	-	-	0.2080	6.1672
b_{n-4}	-	-	-	-	0.1116	3.2802
b_{n-5}	-	-	-	-	0.0960	2.9754
N	9067		9067		9067	
$\ln \mathcal{L}$	-6180.93		-6177.57		-5169.55	
$\ln \mathcal{L}_0$	-6191.96		-6191.96		-6188.88	
LR_0	22.0605		28.7946		2038.657	
$P(LR_0)$	0.0000		0.0000		0.0000	
R_{MZ}^2	0.0039		0.005		0.2869	
R_{AN}^2	0.0024		0.0032		0.1837	
R_{MF}^2	0.0018		0.0023		0.1647	
BIC	12389.20		12400.7		10430.22	

N is the number of observations, $\ln \mathcal{L}$ is the value of the maximized log-likelihood function, $\ln \mathcal{L}_0$ is the value of the log-likelihood function when only a constant is estimated, LR_0 is the likelihood ratio statistic for testing the current model against a specification with constant only, $p(LR_0)$ is the corresponding p-value, R_{MZ}^2 is the value of the McKelvey and Zavoina R^2 , R_{AN}^2 is Aldrich and Nelson's R^2 , and R_{MF}^2 is McFadden's R^2 . t -values have been computed based on QML estimates of the information matrix.

Figures

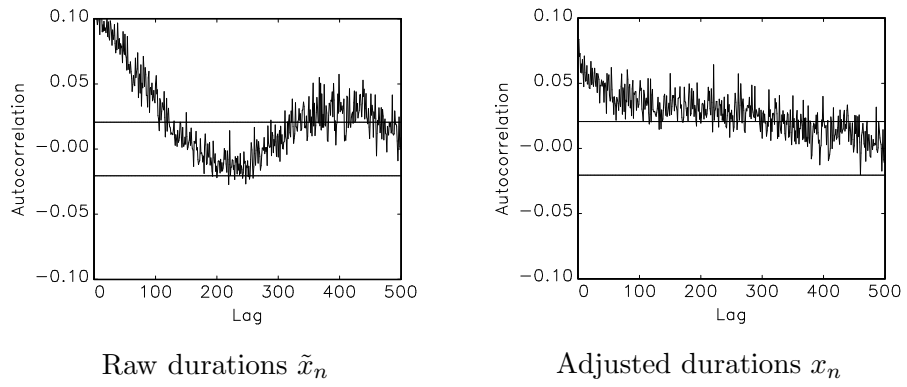


Fig. 1. Autocorrelation function for durations

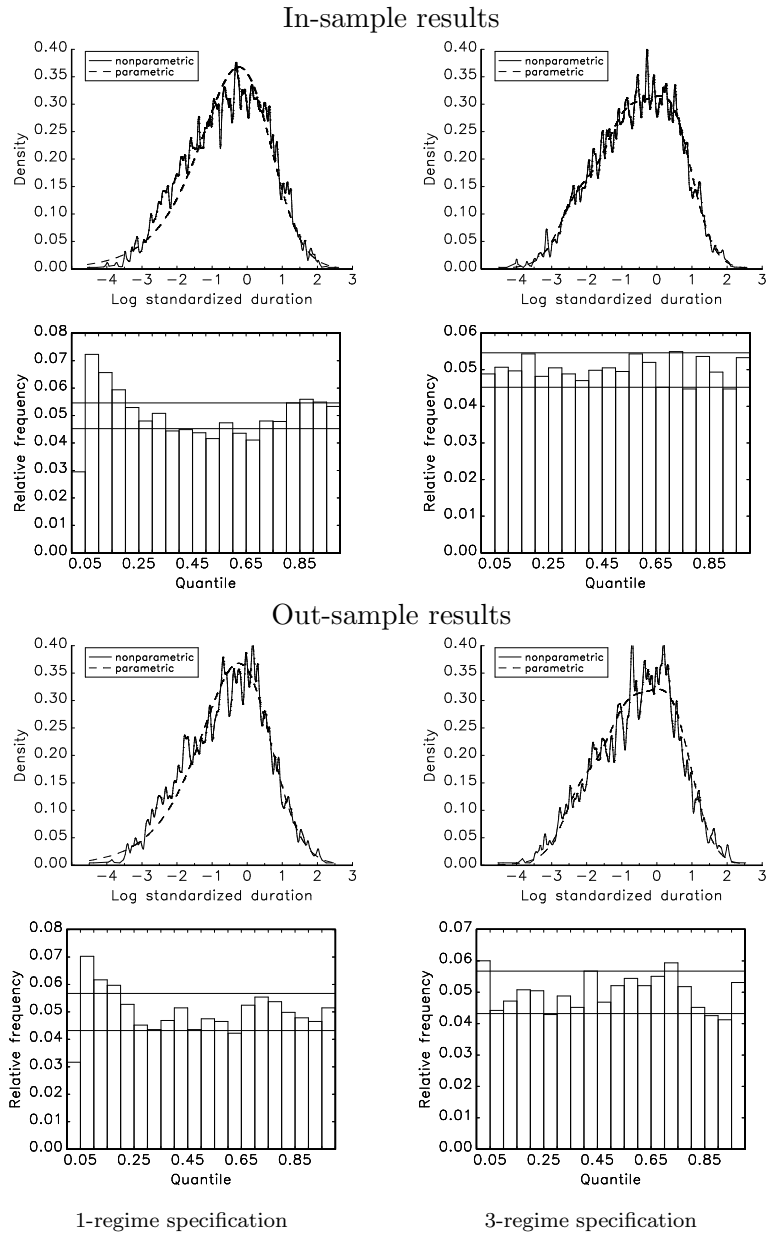


Fig. 2. Results of specification tests. First and third row: Estimates of the density and corresponding theoretical density of log residuals. Second and fourth row: Histogram plots of the ζ_n sequence and 95% confidence intervals.

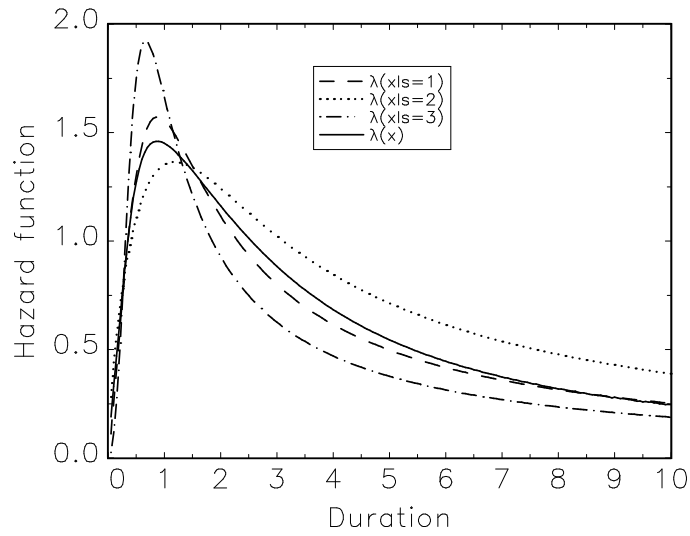


Fig. 3. Regime specific $\lambda(x|s = j)$ and unspecific $\lambda(x)$ hazard function. Evaluated for $\psi_n^{(j)} = 1$ using the total sample estimates of $\kappa^{(j)}$, $\sigma^{(j)}$ and the ergodic probabilities $\pi^{(j)}$ from Table 5.