Hujer, Reinhard; Vuletic, Sandra; Kokot, Stefan

Working Paper

The Markov switching ACD model

Working Paper Series: Finance & Accounting, No. 90

Provided in Cooperation with:
Faculty of Economics and Business Administration, Goethe University Frankfurt

Suggested Citation: Hujer, Reinhard; Vuletic, Sandra; Kokot, Stefan (2002) : The Markov switching ACD model, Working Paper Series: Finance & Accounting, No. 90, Johann Wolfgang Goethe-Universität Frankfurt am Main, Fachbereich Wirtschaftswissenschaften, Frankfurt a. M., https://nbn-resolving.de/urn:nbn:de:hebis:30-18308

This Version is available at:
https://hdl.handle.net/10419/76956

JOHANN WOLFGANG GOETHE-UNIVERSITÄT
FRANKFURT AM MAIN

FACHBEREICH WIRTSCHAFTSWISSENSCHAFTEN

Reinhard Hujer, Sandra Vuletić and
Stefan Kokot

The Markov switching ACD model

No. 90

April 2002

WORKING PAPER SERIES: FINANCE & ACCOUNTING

Reinhard Hujer, Sandra Vuletić and Stefan Kokot

The Markov switching ACD model

No. 90

April 2002

ISSN 1434-3401

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

S. Vuletić
University of Frankfurt, Institute for Statistics and Econometrics, Faculty of Economics and Business Administration, Johann Wolfgang Goethe-University, Mertonstr. 17, 60054 Frankfurt. Phone: +49 69 798 22893 Fax: +49 69 798 23673. Email: vuletic@wiwi.uni-frankfurt.de

S. Kokot
University of Frankfurt, Institute for Statistics and Econometrics, Faculty of Economics and Business Administration, Johann Wolfgang Goethe-University, Mertonstr. 17, 60054 Frankfurt. Phone: +49 69 798 28345 Fax: +49 69 798 23673. Email: kokot@wiwi.uni-frankfurt.de.

Financial support from the Landeszentralbank Hessen gratefully acknowledged. We thank PD Dr. Joachim Grammig for giving us helpful comments and the Gauss codes for the nonparametric specification test. The usual disclaimer applies.

The Markov switching ACD model

Reinhard Hujer, Sandra Vuletić and Stefan Kokot[1,2]

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. Furthermore, we use the MSACD to test implications of a sequential trade model.

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

JEL classification: C22, C25, C41, G14.

[1]Financial support from the Landeszentralbank Hessen is gratefully acknowledged. We thank PD Dr. Joachim Grammig for giving us helpful comments and the Gauss codes for the nonparametric specification test. The usual disclaimer applies.

[2]Faculty of Economics and Business Administration, Johann Wolfgang Goethe-University, Mertonstr. 17, D-60054 Frankfurt.

Contents

1 Introduction

Today it is customary that every single transaction of a financial asset traded on major financial markets around the world is recorded electronically with detailed information about the time of occurrence, price and volume and other relevant characteristics. Recently, many of these *high frequency* data sets have become available at relatively low cost to academic researchers. Hence, the last fifteen years saw an unprecedented upsurge in both theoretical and empirical work related to the analysis of *market microstructure* issues using transaction data sets that are steadily increasing in size.[1] This upsurge went hand in hand with the progress made in computer technology, that made empirical analysis of such data with ordinary desktop computers possible.

It seems only natural then, that this innovation in both, the quality of data available for research and the development of the relevant theory was accompanied by the introduction of new econometric methods which were tailor-made for the analysis of many related policy 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[2] and econometric tools for the analysis of transition data[3] and is therefore perfectly suited for the analysis of high frequency data sets which, unlike most other time series used before in finance and economics, 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 applied to high frequency financial data stemming from stock and foreign exchange markets.

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

[1] See Madhavan (2000) for a recent summary of this branch of literature.
[2] The ACD model is related to the GARCH approach introduced by Engle (1982) and Bollerslev (1986).
[3] See e.g. Lancaster (1990).

several directions.[4] Despite the resulting variety of competing ACD-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.[5] The main problem is the inability of existing ACD models to forecast observations in the tails of their distribution, especially very short trade durations, 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 be achieved by the introduction of an additional latent, discrete valued regime variable whose evolution in time is governed by a *Markov chain*. The unobservable regime can be associated with the presence (or absence) of *private information* about an asset 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 inclusion of latent information structures in an 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 information regime and nests many of the existing ACD models as special cases.

This paper is structured as follows: In Section 2 a brief review of the current state of art in ACD modelling will be given. The MSACD model is introduced in Section 3 and compared to related work on regime switching autoregressive models. We propose two different estimation procedures 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. The usefulness of the MSACD approach

[4]Compare e.g. Bauwens and Giot (1997), Lunde (1999), Russell (1999), Russell and Engle (1999), Grammig and Maurer (2000) and Zhang, Russell, and Tsay (2001).

[5]A comparison of the forecast accuracy of various ACD models with respect to a range of duration processes of interest has been conducted by Bauwens, Giot, Grammig, and Veredas (2000).

for testing 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 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 or level of turnover has been traded.[6]

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

$$E(x_n|\mathcal{F}_n) = \psi_n(\mathcal{F}_n;\theta_\psi) \equiv \psi_n, \tag{1}$$

where $\mathcal{F}_n$ may contain lagged dependent as well as lagged and contemporary exogenous variables, i.e. $\mathcal{F}_n = (x_1, \ldots, x_{n-1}, y_1, \ldots, y_n)$, and θ_ψ is the corresponding set of parameters that determines the conditional mean function. In this framework, 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 i.i.d. with a non-degenerate density function $g(\cdot;\theta_\varepsilon)$ determined by parameters[7] θ_ε and support on the positive real line and an unconditional

[6] 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.

[7] The parameters of the conditional mean θ_ψ and of the conditional density θ_ε are assumed to be *variation free* in the sense of Engle, Hendry, and Richard (1983), i.e. if $\theta_\psi \in \Theta_\psi$ and $\theta_\varepsilon \in \Theta_\varepsilon$, then $\theta \equiv (\theta_\psi, \theta_\varepsilon) \in \Theta_\psi \times \Theta_\varepsilon$.

expectation equal to unity. The flexibility of the ACD model can be altered in (at least) two ways: By modifying the distributional assumption for ε_n and/or the functional form of the conditional mean function $\psi_n(\mathcal{F}_n;\theta_\psi)$.

The choice of $g(\varepsilon_n)$ determines the density of x_n, $f_n\left(x_n \mid \mathcal{F}_n;\theta\right)$ with $\theta=(\theta_\psi,\theta_\varepsilon)$ and will always belong to the same family of distributions as $g(\cdot)$. An assortment of admissible distributions is given in Table VI in the Appendix.

In the most simple case of the ACD model, the parameterization of the conditional mean ψ_n is completely analogous to the parameterization of the conditional variance in a GARCH model. Thus, an ACD(p,q) model arises when the conditional mean function is given by the linear autoregressive specification

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

which 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.

In order to ensure non-negativity of ψ_n, the parameters in the linear specification have to obey some constraints. Alternatively, a nonlinear specification for the conditional mean that closely resembles the EGARCH model of Nelson (1991) can be used. The logarithmic ACD (LACD) specification

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

has been suggested by Bauwens and Giot (1997).[8] Analytical expressions for the unconditional moments of x_n in the LACD specification are quite cumbersome.[9] Imposing stationarity requires additional restrictions on the parameters of the ACD and LACD specification.

Estimates of the parameter vector are most conveniently obtained by maximum likelihood techniques. The log-likelihood function can be expressed either as a sum of the loga-

[8]Several other versions of the LACD specification are obtained by inserting different forms of shocks instead of the lagged durations in (4).

[9]See Bauwens and Giot (2000) for details.

rithms of their conditional densities or by their associated conditional intensities though it is more convenient to estimate θ from the log-likelihood in terms of the duration densities

$$\ln \mathcal{L} = \sum_{n=1}^{N} \ln \left[f_n \left(x_n \mid \mathcal{F}_n; \theta \right) \right], \tag{5}$$

as defined previously. The estimation results can be used to examine if an acceptable specification for the dynamics of the duration process has been found. Many tests are based on the distributional assumption for ε_n.

3 The Markov switching ACD model

3.1 Regime switching models in econometrics

Regression models that allow for time variable conditional mean specifications have a long history in econometrics. Apart from the literature on testing for structural changes (e.g. Chow (1960), Goldfeldt 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. Switching regression models first appeared in Goldfeldt and Quandt (1973). In these models, 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 is itself an autoregressive process, whose behavior is governed by a hidden Markov chain - which itself could depend on observable variables.

In a seminal paper, Hamilton (1989) 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, (Turner, Startz, and Nelson (1989), Engel and Hamilton (1990), Cecchetti, Lam, and Mark (1990), Mundaca (2000), and Dewachter (2001)). The common link between the MSAR model and the earlier literature on static switching regression models is that both approaches imply that the data generating process of the dependent variable can be described by a discrete mixture density, where the conditional

density of the dependent variable, given the regime, is specified to be from some known family of distributions, usually the Gaussian, and the density of the regime variable is left unspecified. The regime probabilities are estimated non-parametrically along with the regression parameters, by imposing the restriction, that the regime density is discrete valued and has a known, finite number of support points. The MSAR model has been extended in many ways, e.g. by allowing for time-varying transition probabilities in the markov chain (Filardo (1994), Gray (1996), and Diebold, Lee, and Weinbach (1997)) or to model changes in the conditional variances in an ARCH model (Cai (1994), Hamilton and Susmel (1994), Gray (1996), and Rydén, Teräsvirta, and Åsbrink (1998)).

Mixture models for point processes have also a long history. In the context of event history analysis, mixture models have often been used to model unobserved individual heterogenity in cross-sectional or panel data (Heckman and Singer (1982), Gritz (1993) and Allenby, Leone, and Jen (1999)). There are also many applications of mixture models for count data, (Wedel, Desarbo, Bult, and Ramaswamy (1993), Wang, Puterman, Cockburn, and Le (1996), Trivedi and Deb (1997) or Wang, Cockburn, and Puterman (1998)), but none of these papers treats the case of a time series of events. A notable exception is the material summarized in the book by MacDonald and Zucchini (1997), who motivate hidden Markov models as an alternative to other time series models for discrete valued processes. They avoid to generalize their analysis to point process modeling. Also, they do not treat autoregressive specifications of the conditional mean of the observable time series.

3.2 The MSACD model

We will apply the Markov switching framework as a statistical model for financial duration processes. 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 or state the process is in at time t_n. At the next time period, the duration process can switch to another regime. 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 given the filtration $\mathcal{F}_n$ the decomposition (2) holds in the sense that $E(\varepsilon_n \mid \mathcal{F}_n) = 1$. The conditional mean of the duration x_n depends on the unobserved regime variable s_n in the following manner

$$\psi_n = \sum_{j=1}^{J} p(s_n = j \mid \mathcal{F}_n; \theta) \cdot \psi_n^{(j)}, \tag{6}$$

where $p(s_n = j \mid \mathcal{F}_n; \theta)$ is the probability, that s_n is in state j given the filtration $\mathcal{F}_n$. The regime specific conditional mean $\psi_n^{(j)} = E(x_n \mid s_n = j, \mathcal{F}_n; \theta)$ depends on an associated set of parameters $\theta_\psi^{(j)}$ and may have an autoregressive specification as in Section 2.

The regime variable s_n switches between the states according to a Markov chain which is characterized by a transition matrix $\mathbf{P}$ 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 $f_n(x_n \mid s_n = j, \mathcal{F}_n; \theta)$ depends only on the current regime s_n and on $\mathcal{F}_n$ where $\theta \in \Theta$ is a vector containing all parameters that describe the distribution of x_n, $\theta = \left(\theta_\varepsilon^{(1)}, \ldots, \theta_\varepsilon^{(J)}, \theta_\psi^{(1)}, \ldots, \theta_\psi^{(J)}, p_{11}, \ldots, p_{JJ}\right)'$. Any of the densities we introduced in Table VI 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 of the observed durations[10]

$$f_n(x_n \mid \mathcal{F}_n; \theta) = \sum_{j=1}^{J} p(s_n = j \mid \mathcal{F}_n; \theta) \cdot f_n(x_n \mid s_n = j, \mathcal{F}_n; \theta). \tag{7}$$

In order to evaluate this marginal density in a Markov switching model, the *filtered* regime probability

$$\xi_{n+1|n}^{(j)} \equiv p(s_{n+1} = j \mid \mathcal{F}_{n+1}; \theta), \tag{8}$$

plays an 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

[10] In the context of the EM-algorithm introduced in Section 3.5, this density is also referred to as the *incomplete data* density.

obtained from a two-step recursion as follows[11]

$$\xi_{n|n}^{(j)} = \frac{\xi_{n|n-1}^{(j)} \cdot f_n\left(x_n \mid s_n = j, \mathcal{F}_n; \theta\right)}{\sum_{k=1}^{J} \xi_{n|n-1}^{(k)} \cdot f_n\left(x_n \mid s_n = k, \mathcal{F}_n; \theta\right)} \tag{9}$$

$$\xi_{n+1|n}^{(j)} = \sum_{i=1}^{J} p_{ji} \cdot \xi_{n|n}^{(i)}, \tag{10}$$

With a given set of start values $\xi_{1|0}^{(j)}$ 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 $\xi_{n|n}^{(j)}$ and $\xi_{n+1|n}^{(j)}$ 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 all equal, i.e. $\pi_j \equiv p_{j1} = \ldots = p_{jJ}$. 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. This is the key property that distinguishes a static mixture model from the Markov switching model.

An issue that has to be addressed, is the specification of the conditional mean function $\psi_n^{(j)}$. There are in principle two possible ways, in which lagged forecasts can appear. If the current forecast $\psi_n^{(j)}$ is a function of $\psi_{n-1}^{(j)}, \psi_{n-2}^{(j)}, \ldots$, forecasts are based on the regime specific forecast error. Note, that the regime specific conditional mean of an MSACD(1,1) model admits the ARMA(1,1) representation $x_i = \omega^{(j)} + (\alpha^{(j)} + \beta^{(j)}) \cdot x_{i-1} - \beta^{(j)} \cdot (x_{i-1} - \psi_{i-1}^{(j)})$, thus the model implicitly assumes, that economic agents revise their forecasts based on the deviation between the realized duration x_{i-1} and the regime specific expectation $\psi_{i-1}^{(j)}$. Another possible specification is to make $\psi_n^{(j)}$ a function of past forecasts that are regime independent ψ_{n-1}, thus assuming that the deviation between the realized duration and the unconditional expectation $(x_{i-1} - \psi_{i-1})$ determines the current regime forecast.

However, when regime independent lagged expectations ψ 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

[11] See Hamilton (1994), p. 692-694 for a proof.

all of the possible paths even for a moderate sample size is prohibitively expensive in terms of computational effort. 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 $\psi_n^{(j)}$

$$\psi_n = \sum_{j=1}^{J} \xi_{n|n-1}^{(j)} \cdot \psi_n^{(j)}. \tag{11}$$

When this specification for the conditional mean is employed, the distribution of $\varepsilon_n = \frac{x_n}{\psi_n}$ can be derived as shown in Appendix A.2.

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 heterogenity 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 (9) 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* $\xi_{n|N}^{(j)} = p(s_n = j \mid x_N, \mathcal{F}_N; \theta)$. These may be evaluated using an algorithm, which consists of a backward recursion starting with the filtered inference $\xi_{N|N}^{(j)}$ obtained from (9) and progressing according to

$$\xi_{n|N}^{(j)} = \xi_{n|n}^{(j)} \cdot \sum_{k=1}^{J} \frac{p_{kj} \cdot \xi_{n+1|N}^{(k)}}{\xi_{n+1|n}^{(k)}}. \tag{12}$$

This algorithm has been proposed by Kim (1994). 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 $\mathcal{F}_n$.

3.4 ML-estimation of MSACD models

In the case of regime switching models there are several ways, in which ML 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

$$\ln \mathcal{L}_I(\theta) = \sum_{n=1}^{N} \ln \left[f_n(x_n \mid \mathcal{F}_n; \theta) \right], \tag{13}$$

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 non-negativity, stationarity and eventually for distributional parameters. If the conditional mean function $\psi_n^{(j)}$ contains lagged expectations ψ_{n-k}, the likelihood function has to be evaluated observation by observation for all regimes simultaneously, because of the need to aggregate lagged conditional means $\psi_{n-k}^{(j)}$ according to equation (11). Also, filtered inferences on the regime probabilities given in equations (9) and (10) have to be evaluated observation by observation. Since this has to be repeated several times during each iteration, the resulting estimation procedure is very time consuming.

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 MSACD models

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

$$\begin{aligned} \ln \mathcal{L}_C(\theta) &= \ln \left[\prod_{n=1}^{N} f_n(x_n, s_n \mid \mathcal{F}_n; \theta) \right] \\ &= \sum_{n=1}^{N} \sum_{j=1}^{J} z_n^{(j)} \cdot \ln[f_n(x_n \mid s_n = j, \mathcal{F}_n; \theta)] + \sum_{n=2}^{N} \sum_{j=1}^{J} \sum_{i=1}^{J} z_n^{(ji)} \cdot \ln[p_{ji}]. \end{aligned} \tag{14}$$

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. The EM-algorithm proceeds by taking the expectation of (14) conditional on the observable data $\mathcal{X}_N = (x_1, \ldots, x_N, y_1, \ldots, y_N)$ and evaluates it using some preliminary guess for the parameter vector θ_0. The *expected complete* log-likelihood function $\ln \mathcal{L}_{EC} \equiv E[\ln \mathcal{L}_C(\theta) \mid \mathcal{X}_N; \theta_0]$ for the MSACD is given by

$$\begin{aligned} \ln \mathcal{L}_{EC}(\theta, \theta_0) &= \sum_{n=1}^{N} \left(\sum_{j=1}^{J} \xi_{n|N}^{(j)} \cdot \ln \left[f_n(x_n \mid s_n = j, \mathcal{F}_n; \theta) \right] \right) \\ &\quad + \sum_{n=2}^{N} \left(\sum_{j=1}^{J} \sum_{i=1}^{J} \xi_{n|N}^{(ji)} \cdot \ln \left[p_{ji} \right] \right), \end{aligned} \tag{15}$$

where $\xi_{n|N}^{(j)}$ is the full sample inference on the regime obtained by evaluating the backward recursion (12) using the preliminary guess θ_0 and[12]

$$\xi_{n|N}^{(ji)} = p\left(s_n = j, s_{n-1} = i \mid x_N, \mathcal{F}_N; \theta_0\right) = \xi_{n-1|n-1}^{(i)} \cdot \frac{p_{ji} \cdot \xi_{n|N}^{(j)}}{\xi_{n|n-1}^{(j)}}. \tag{16}$$

[12] See Hamilton (1994), p. 701.

The 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*. 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 $\mathcal{X}_N$ and evaluated using an arbitrary guess for the parameter vector θ_0.

The associated *M-step* consists of maximizing the expected complete log-likelihood function $\ln \mathcal{L}_{EC}(\theta, \theta_0)$ with respect to the parameter vector θ. The same restrictions as in the case of the incomplete log-likelihood have to be imposed. 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 $\frac{\partial f_n(x_n|s_n=j,\mathcal{F}_n;\theta)}{\partial p_{mk}} = 0, \ \forall \ j, m, k \in (1, \ldots J)$. Then the first order conditions lead to the following estimator for the transition probabilities[13]

$$\hat{p}_{ji} = \frac{\sum_{n=2}^{N} p\left(s_n = j, s_{n-1} = i \mid x_N, \mathcal{F}_N; \theta_0\right)}{\sum_{n=2}^{N} p\left(s_{n-1} = i \mid x_N, \mathcal{F}_N; \theta_0\right)} = \frac{\sum_{n=2}^{N} \xi_{n|N}^{(ji)}}{\sum_{n=2}^{N} \xi_{n-1|N}^{(i)}}, \tag{17}$$

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. The remaining parameters may be obtained from the solution to

$$\sum_{n=1}^{N} \sum_{j=1}^{J} \xi_{n|N}^{(j)} \cdot \left(\frac{\partial \ln f_n(x_n \mid s_n = j, \mathcal{F}_n; \theta)}{\partial \theta} \right) \stackrel{!}{=} 0. \tag{18}$$

Note that the condition $\frac{\partial f_n(x_n|s_n=j,\mathcal{F}_n;\theta)}{\partial p_{mk}} = 0, \ \forall \ j, m, k \in (1, \ldots J)$ is satisfied when (a) the specification of the conditional mean function of the ACD-part of the model does not contain lagged expected durations (so the aggregation procedure described in Section 3.2 can be avoided), (b) the lagged expected durations appearing in the conditional mean specification are taken to be dependent on the same regime as the current forecast as e.g. in $\psi_n^{(j)} = \omega^{(j)} + \beta^{(j)} \cdot \psi_{n-1}^{(j)} + \alpha^{(j)} \cdot x_{n-1}$ or (c) only distributional parameters θ_ε are regime dependent.

[13] See Hamilton (1989).

The EM-algorithm proceeds iteratively in the following manner: Each iteration of the algorithm contains two steps. Starting from an initial guess for the parameter vector θ_0, the E-step consists of the computation of the smoothed probabilities $\xi_{n|N}^{(j)}$ and $\xi_{n|N}^{(ji)}$. These probabilities are plugged into the formula for the expected complete log-likelihood function. In the M-step, this function is being maximized in order to obtain an updated guess for the parameter vector $\hat{\theta_1}$ which will be used to conduct the E-step in the next iteration of the algorithm. Thus, by repeating these two steps until a prespecified convergence criterion 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.[14]

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 non-standard 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. The corresponding likelihood ratio statistic will not have the usual χ^2- distribution, but differ from it substantially even in large samples.[15] Another example is the usual t-statistic for $H_0 : p_{ji} = 0$ against $H_A : p_{ji} > 0$. Under the Null, 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

[14] See Hamilton (1990) for a proof.
[15] See Böhning, Dietz, Schaub, Schlattman, and Lindsay (1994).

critical values of the t-distribution. In the context of Markov-switching regression models, Hamilton (1996) proposes a variety of specification tests, which are based either on conditional moment conditions implied by the model or on the Lagrange multiplier principle. Both types of tests require the evaluation of the score function $l_n = \frac{\partial \ln f_n(x_n|\mathcal{F}_n;\theta)}{\partial \theta}$, where $f_n(x_n|\mathcal{F}_n;\theta)$ is the density defined in (7). Hamilton derives analytical expressions for the scores of the incomplete log-likelihood function $\ln \mathcal{L}_I(\theta)$ and all test statistics he considers are functions of the scores.

Fernandes and Grammig (2000) have introduced specification tests for ordinary ACD models which are based on the discrepancy between the observed and the theoretical density respectively hazard function of the residuals and are, with minor refinements, applicable to the MSACD model as well. In the following, we will discuss the test based on the density of the residuals only. In ordinary ACD models the test statistic is easily calculated by noting that the residuals ε are independently identically distributed (i.i.d.). The null hypothesis is

$$H_0 : \exists \quad \theta \in \Theta \quad \text{such that} \quad g(\varepsilon;\theta) = g(\varepsilon) \tag{19}$$

where $g(\varepsilon)$ is the true but unknown density of the residuals and $g(\varepsilon;\theta)$ the density implied by the parametric ACD 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 from an ACD specification $g(\varepsilon;\hat{\theta})$. Thus the observed mean squared distance D_g between the two densities is given by

$$D_g = \frac{1}{N}\sum_{n=1}^{N}\left[g(\hat{\varepsilon}_n;\hat{\theta}) - \hat{g}(\hat{\varepsilon}_n)\right]^2. \tag{20}$$

Under the null hypothesis (19) and some additional regularity conditions[16] the statistic FG has asymptotically a standard normal distribution. FG is given by

$$FG = \frac{N \cdot h^{0.5} \cdot D_g - h^{-0.5} \cdot \hat{E}_{D_g}}{\sqrt{\hat{V}_{D_g}}}, \tag{21}$$

[16] For details, see Fernandes and Grammig (2000).

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[17], $\hat{E}_{D_g}$ and $\hat{V}_{D_g}$ are consistent estimates of

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

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.

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.[18] The Fernandes and Grammig (2000) test can still be conducted, though one has to take the time varying nature of the theoretical density into account. The kernel density estimator provides us with an estimate of the density $g(\varepsilon)$, which is itself a mixture of the N different densities $g_n(\varepsilon \mid \mathcal{F}_n; \theta)$. Therefore we propose to make the two densities comparable by taking the mean of the theoretical densities

$$g(\varepsilon; \hat{\theta}) = \frac{1}{N} \cdot \sum_{n=1}^{N} g_n(\varepsilon \mid \mathcal{F}_n; \hat{\theta}), \tag{22}$$

and calculate the distance to the kernel density estimate.

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

[17] 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. E.g. 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.

[18] See Appendix A.2.

density will be preferred by all forecast users regardless of the form of their loss functions. This suggests that forecast performance can be evaluated by assessing whether the densitiy of the forecasts are correct, i.e. whether

$$H_0 : \quad \{f_n(x_n \mid \mathcal{F}_n; \hat{\theta})\}_{n=1}^N = \{f_n(x_n \mid \mathcal{F}_n; \theta)\}_{n=1}^N. \tag{23}$$

Since the true distribution $f_n(x_n \mid \mathcal{F}_n; \theta)$ is never observed, the sequence of conditional empirical distribution functions defined by

$$\hat{\zeta}_n = \int_{-\infty}^{x_n} f_n(u \mid \mathcal{F}_n; \hat{\theta})\, du \tag{24}$$

is used as a test statistic. As shown by Rosenblatt (1952), under the null hypothesis the distribution of the sequence of probability transforms $\hat{\zeta}_n$ is uniform i.i.d. on the unit interval, so that any test for uniformity of the $\hat{\zeta}_n$ sequence can be used to asses the forecast performance of the model under consideration. The test can also be used to compare the performance of several non-nested model specifications.

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 $\hat{\zeta}_n$ sequence. A straightforward χ^2 goodness-of-fit test can be conducted by exploiting the statistical properties of the histogram under the null hypothesis of uniformity.

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) during a month in 1996 from the trades and quotes database (TAQ) provided by the NYSE Inc. The sampling period spans 19 trading days from Nov. 1 until Nov. 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.

4.2 Analysis of seasonality

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, see Figure 1. Engle and Russell (1997) propose to decompose the duration series into a deterministic function of the time of day, $\Phi(t_n)$ and a stochastic component x_n, so that $\tilde{x}_n = x_n \cdot \Phi(t_n)$. In this paper, we will apply the two step method proposed by Engle and Russell (1997) in which the time of day function is estimated separately from the other model parameters.[19] Dividing each duration in the sample by the appropriate time of day function value, a sequence of deseasonalized durations is obtained.

In order to estimate the time of day function we employ the semi-nonparametric (SNP) estimator introduced by Gallant (1981) and Eubank and Speckman (1990). The basic approach is to approximate the unknown function $\Phi(t_n)$ using a fourier series expansion accommodated by polynomials in the regressor variables. Estimation in the univariate case is carried out by fitting a regression curve of the type

$$\tilde{x}_n = \beta_0 + \sum_{p=1}^{P} \left(\beta_p \cdot h(t_{n-1})^p\right) + \sum_{q=1}^{Q} \left[\phi_q \cdot \cos\left(q \cdot h(t_{n-1})\right) + \delta_q \cdot \sin\left(q \cdot h(t_{n-1})\right)\right], \tag{25}$$

where the normalizing function $h(t)$ is given by

$$h\left(t\right) = 2\pi \cdot \frac{t - t_{\min}}{t_{\max} - t_{\min}}, \tag{26}$$

[19] Simultaneous ML estimation would also be possible. Engle and Russell (1998) report that both procedures give similar results if sufficient data is available.

and $t_{\min}$ ($t_{\max}$) is the time of day at which trading begins (ends) at the NYSE. This type of estimator is especially well suited for our purposes, since it can reproduce non-linear shapes of the time of day function.[20]

Our choice of the smoothing parameters P and Q was guided by the analysis in Andrews (1991b), who shows that when the errors of the regression function are heteroscedastic, smoothing parameter selection by minimizing the simple 'leave one out' cross-validation function over a grid of values for P and Q will lead to asymptotically optimal estimates. Thus we re-estimated the model for the grid defined by $P = 1, 2, 3$ and $Q = 1, \ldots, 25$ and found a minimum of the cross-validation function at $P = 1$ and $Q = 1$. A plot of the cross-validation function and the corresponding estimated intraday pattern is contained in the right panel of Figure 1. In order to assess the fit of the estimated seasonal pattern, we included the intradaily evolution of the mean durations, computed over successive 5 minute intervals in the left panel of Figure 1. This alternative estimator of the seasonal pattern is clearly more severely affected by sampling variation than the SNP-estimate, and also has the drawback to be a discontinuous function.

Descriptive information about sample moments and Ljung Box statistics of the original and the seasonally adjusted duration data are reported in Table I. 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. Even after seasonal adjustment, the Ljung-Box tests for no autocorrelation up to 50 lags are rejected at the 5% significance level. 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 9092 deseasonalized durations into two subperiods. The column

[20] Also the SNP approach takes into account that the regressor variable has bounded support, which is true in our application where the trading day is limited to 6.5 hours per day. Asymptotic normality and consistency of SNP-estimators for several types of data generating processes with i.i.d. and heteroscedastic errors have been established in Eastwood (1991), and Andrews (1991a). The same technique has been applied by several researchers to estimate seasonal components in GARCH-models, see e.g. Andersen and Bollerslev (1997).

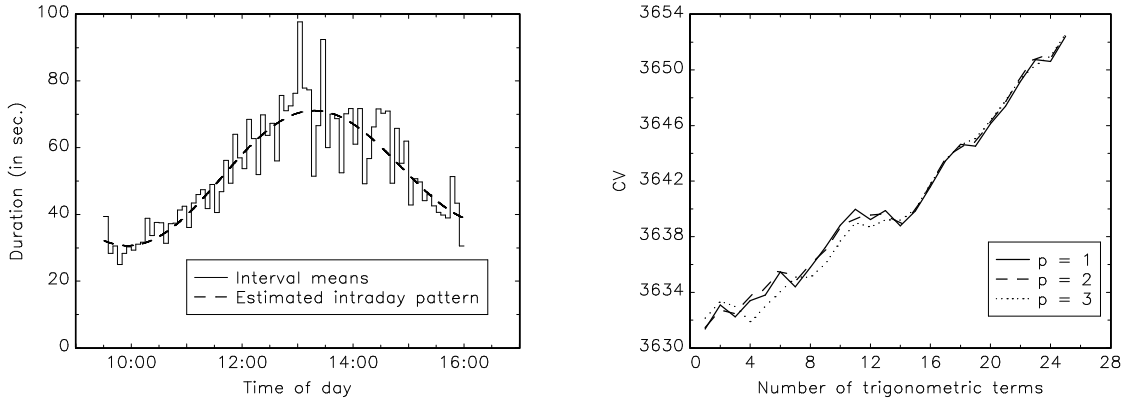

FIGURE 1: Analysis of intraday seasonality. Left panel: Intradaily 5 minute interval means and estimated seasonal pattern. Right panel: Cross-validation function.

titled "In-sample" contains the descriptive statistics for the first 6060 observations (corresponding to two thirds of the total sample), which are used 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 characteristics, except for the occurrence of very large durations, which tend to appear more concentrated in the first subsample. This could explain the higher excess kurtosis as well as the higher value of the Ljung Box statistic in the first subsample. It is well known, that estimates of the autocorrelation function are sensible to extreme observations (outliers). Nevertheless, note that although the Ljung Box statistic is much smaller for the second subsample than for the first, it is still consistent with rejection of the null hypothesis of no autocorrelation at conventional significance levels.

4.3 Estimation results for MSACD models

We focus on the logarithmic class of ACD models based on the Burr distribution for each regime and distinguish between three different specifications of the MSACD model. In the *restricted* (R) MSACD model specification, the restrictions $\omega^{(1)} = \ldots = \omega^{(J)}$, $\beta^{(1)} = \ldots = \beta^{(J)}$, $\alpha^{(1)} = \ldots = \alpha^{(J)}$ and $p_{j1} = \ldots = p_{jJ} \equiv \pi_j$ have been imposed, so that only

TABLE I: Descriptive Statistics for Boeing 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
Overdispersion	79.1392	1.4229	1.4905	1.2575
Skewness	3.3137	2.8681	2.8944	2.6647
Excess Kurtosis	18.8846	13.5436	14.2189	9.4243
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.

the remaining distributional parameters are allowed to vary between regimes. The *simple unrestricted* (SUR) model is characterized by the feature that lags of the regime specific conditional mean may appear in the forecast function, yielding e.g. the following first order specification $\ln(\psi_n^{(j)}) = \omega^{(j)} + \beta^{(j)} \cdot \ln(\psi_{n-1}^{(j)}) + \alpha^{(j)} \cdot \ln(x_{n-1})$. Estimation of the R and SUR specification can exploit the advantages of the EM-algorithm, while the *unrestricted* (UR) MSACD model, which includes lags of the regime independent conditional means obtained by aggregation of regime specific means in the forecast function as in $\ln(\psi_n^{(j)}) = \omega^{(j)} + \beta^{(j)} \cdot \ln(\psi_{n-1}) + \alpha^{(j)} \cdot \ln(x_{n-1})$, has to be estimated by maximization of the incomplete log-likelihood function.

The in-sample results of the specification tests for all of the model specifications we estimated are presented in Table II. None of the specification tests that we performed, supports the one regime model. Also, the Bayesian information criterion[21] (BIC) does

[21] See Schwarz (1978).

TABLE II: Specification tests for Burr MSACD models, in-sample

Model	$\ln \mathcal{L}$	BIC	$P(\chi^2_\zeta)$	$P(LB_\zeta)$	$P(FG)$	$P(LB_\varepsilon)$
			1 Regime specification			
$LACD(0,1)$	-6166.24	12367.33	0.0000	0.0000	0.0000	0.0000
$LACD(1,1)$	-6025.59	12094.74	0.0000	0.3335	0.0000	0.6315
			2 Regime specification			
$LACD(0,1)$						
R	-6021.99	12104.94	0.0288	0.0000	0.0000	0.0000
SUR/UR	-6001.59	12090.28	0.0000	0.1385	0.0000	0.0000
$LACD(1,1)$						
R	-5898.96	11867.60	0.0912	0.4165	0.0053	0.4100
SUR	-5864.46	11833.44	0.1145	0.4048	0.0015	0.4905
UR	-5883.20	11870.92	0.0125	0.2562	0.0020	0.5505
			3 Regime specification			
$LACD(0,1)$						
R	-6021.99	12131.07	0.0288	0.0000	0.0000	0.0000
SUR/UR	-5899.55	11955.87	0.0172	0.0137	0.0000	0.0000
$LACD(1,1)$						
R	-5898.96	11893.73	0.0823	0.3757	0.0056	0.3728
SUR	-5828.55	11840.00	0.2666	0.1609	0.0581	0.1822
UR	-5840.06	11863.03	0.8391	0.2103	0.4758	0.7746
			4 Regime specification			
$LACD(0,1)$						
R	-6000.27	12113.77	0.9074	0.0000	0.2248	0.0000
SUR/UR	-5869.27	11982.42	0.1757	0.0019	0.0337	0.0000
$LACD(1,1)$						
R	-5898.96	11919.85	0.0823	0.3369	0.0056	0.3340
SUR	-5804.04	11886.78	0.4558	0.1188	0.7985	0.5552
UR	-5834.64	11947.99	0.0556	0.0194	0.0349	0.1278

$\ln \mathcal{L}$ is the value of the log-likelihood function, BIC is the Bayesian information criterion, computed as $-2 \cdot \ln \mathcal{L} + \ln(N) \cdot k$, where k is the number of estimated parameters, $P(\chi^2_\zeta)$ is the p-value of the ordinary χ^2 goodness of fit statistic for the i.i.d. uniformity of ζ, using an histogram estimator for its density based on 20 bins, $P(LB_\zeta)$ is the p-value corresponding to the Ljung-Box statistic for 50 lags of ζ, $P(FG)$ is the p-value of the 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.

not support this model. From the plots of the density estimates of the residuals, as well as from the histogram of the series of density integral transforms ζ_n (see Figure 2), we find that one regime ACD-models have severe problems to predict very small durations appropriately. These findings also hold for the out-sample forecasting performance of the

one regime models. Neither the FG nor the uniformity test for the ζ_n series is passed at conventional significance levels, as indicated by the p-values of the corresponding test statistics given in Table III. Also, the one regime models perform bad in terms of forecast accuracy, as indicated by the values of the mean squared error (MSE) and mean absolute error (MAE).

TABLE III: Specification tests for Burr MSACD models, out-sample

Model	$P(\chi^2_\zeta)$	$P(LB_\zeta)$	$P(FG)$	$P(LB_\varepsilon)$	MSE	MAE
			1 Regime specification			
$LACD(0,1)$	0.0000	0.0000	0.0039	0.0000	1.1687	0.7782
$LACD(1,1)$	0.0000	0.3802	0.0053	0.4840	1.1268	0.7268
			2 Regime specification			
$LACD(0,1)$						
R	0.0000	0.0000	0.0000	0.0000	1.2998	0.9006
SUR/UR	0.0002	0.3464	0.4155	0.2602	1.1366	0.7486
$LACD(1,1)$						
R	0.1278	0.4022	0.0065	0.3427	1.1976	0.8200
SUR	0.0228	0.2724	0.0011	0.3382	1.1255	0.7252
UR	0.0296	0.2762	0.0002	0.3258	1.1271	0.7266
			3 Regime specification			
$LACD(0,1)$						
R	0.0000	0.0000	0.0000	0.0000	1.2999	0.9006
SUR/UR	0.2658	0.1309	0.6546	0.0490	1.1441	0.7564
$LACD(1,1)$						
R	0.1337	0.3627	0.0065	0.3057	1.1976	0.8200
SUR	0.0124	0.0930	0.0000	0.1979	1.1227	0.7205
UR	0.0008	0.1466	0.0000	0.1527	1.1236	0.7271
			4 Regime specification			
$LACD(0,1)$						
R	0.0000	0.0000	0.0000	0.0000	1.3696	0.9494
SUR/UR	0.0063	0.0269	0.3265	0.0162	1.1356	0.7480
$LACD(1,1)$						
R	0.1337	0.3243	0.0065	0.2704	1.1976	0.8200
SUR	0.0105	0.0156	0.0416	0.0640	1.1247	0.7192
UR	0.0008	0.0217	0.0003	0.0127	1.1257	0.7302

$MSE = N^{-1} \cdot \sum_{n=1}^{N} (x_n - \hat{\psi}_n)^2$, $MAE = N^{-1} \cdot \sum_{n=1}^{N} |x_n - \hat{\psi}_n|$.

Turning to the MSACD model, we estimated 2-, 3-, and 4-regime specifctions with LACD(0,1) and LACD(1,1) conditional mean functions. 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. With regard to the in-sample results, the SUR MSACD specification performed generally better than the corresponding R and UR versions, both in terms of the values of the BIC and the results of the specification tests that we conducted. The BIC prefers the 2-regime specification, but the results of the FG tests do not support the 2-regime specification at all. Therefore, we focused on the 3-regime SUR specification, 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 the MSE and MAE. Focusing on the out of sample results, the SUR specifications tend to give more accurate forecasts than the corresponding U and UR specifications, as indicated by the values of the MSE and MAE. On the other hand, the test statistics for the FG and the uniformity test of the forecast density do seem to favor the R model in the LACD(1,1) case, while the SUR/UR model is preferred in the LACD(0,1) case. Perhaps surprisingly, the models that pass the specification tests tend to perform worse in terms of forecast accuracy than the models that do not pass them.

Even though the result of the FG test for the out-sample does not support the 3-regime SUR LACD(1,1) specification, we find that it offers a reasonable compromise between 'in-sample' and 'out-sample' performance and therefore will be used to conduct tests of the implications of market microstructure theories in Section 5. Table IV contains the corresponding parameter estimates, standard errors and results of the specification tests described above for the selected 3-regime SUR LACD(1,1) specification. 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). The same model was also estimated based on the total sample of 9092 observations for use in Section 5. The parameter estimates for the

entire sample are presented in the column entitled "Total sample" and 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.

TABLE IV: Estimation Results for the *SUR* 3-Regime log. Burr MSACD(1,1) model

	In-sample		Total sample	
Parameter	Estimate	Stderr.	Estimate	Stderr.
$\omega^{(1)}$	-0.0102	0.0107	-0.0165	0.0166
$\alpha_1^{(1)}$	0.0215	0.0072	0.0262	0.0125
$\beta_1^{(1)}$	0.9663	0.0154	0.9511	0.0296
$\kappa^{(1)}$	2.1565	0.1570	2.1955	0.1228
$\sigma^{2(1)}$	0.8749	0.1366	0.8665	0.1286
$\omega^{(2)}$	0.0243	0.0161	0.0187	0.0043
$\alpha_1^{(2)}$	0.0238	0.0159	0.0188	0.0046
$\beta_1^{(2)}$	0.9757	0.0176	0.9812	0.0046
$\kappa^{(2)}$	1.6385	0.0776	1.6577	0.0600
$\sigma^{2(2)}$	0.3559	0.0516	0.4080	0.0444
$\omega^{(3)}$	-0.0181	0.0053	-0.0399	0.0167
$\alpha_1^{(3)}$	0.0031	0.0018	0.0077	0.0043
$\beta_1^{(3)}$	0.9883	0.0029	0.9741	0.0117
$\kappa^{(3)}$	3.4156	0.3910	3.0937	0.2745
$\sigma^{2(3)}$	1.8981	0.3673	1.6414	0.2246
p_{11}	0.4112	0.0701	0.3889	0.0581
p_{12}	0.2572	0.0386	0.2211	0.0341
p_{13}	0.3232	0.0817	0.3279	0.0660
p_{21}	0.4206	0.0821	0.4290	0.0607
p_{22}	0.5932	0.0377	0.6065	0.0382
p_{23}	0.5416	0.0943	0.5410	0.0849
N	6060		9092	
$\ln \mathcal{L}$	-5828.55		-8430.96	
BIC	11840.00		17053.34	
χ^2_ζ	22.3630	***	19.3242	***
LB_ζ	46.5509	***	44.9270	***
FG	1.5708	**	-0.2250	***
LB_ε	45.7174	***	54.4470	*

* indicates that the Null-hypothesis of the corresponding specification test may not be rejected at the 1%-significance level, ** at the 5%-significance level, and *** at the 10%-significance level.

For purposes of comparison, Figure 2 contains plots of the density estimates for $\hat{\varepsilon}$, as well as the histograms for the ζ series for the 1-regime and the 3-regime specifications. 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 wiggly and occasionally lie clearly outside of the confidence interval. Even so, there is no sign of a systematical pattern of over- or underestimation as in the case of the one regime model.

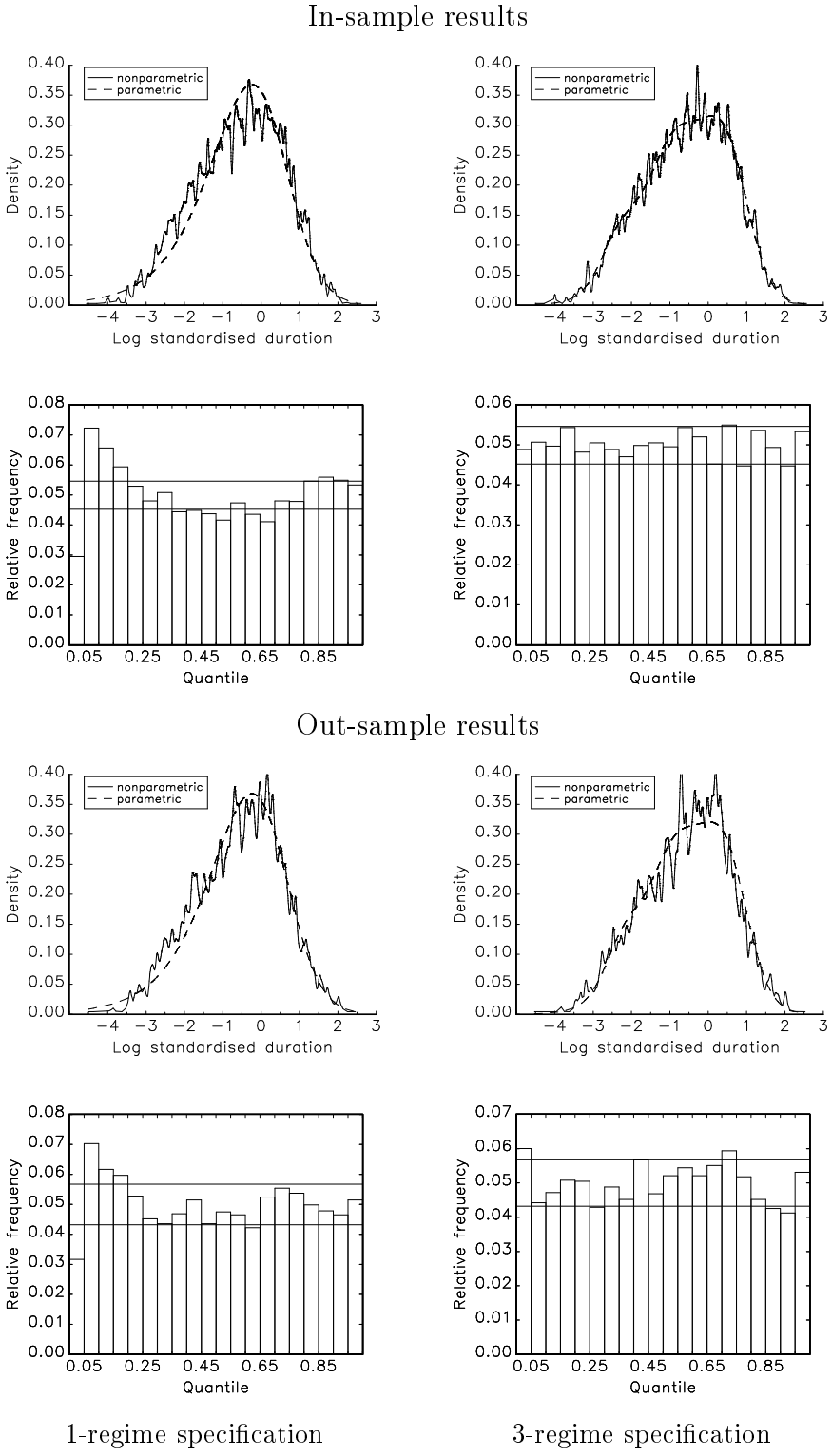

FIGURE 2: Results of the specification tests for 1-regime versus 3-regime MSACD(1,1) models. First row: Estimates of the density of the log residuals and corresponding theoretical density of log residuals implied by the estimated in-sample model. Second row: Histogram plots of the cumulative forecast density and 95% confidence intervals for the in-sample. Third row: Theoretical and estimated density of log residuals for out-sample. Last row: Histogram plot for the out-sample.

5 Testing implications of sequential trade models

In the framework of Easley, Kiefer, O'Hara, and Paperman (1996), henceforth denoted as EKOP, the price setting behavior of market makers is explained by the presence of traders who have superior information affecting future price movements. Their setup is a mixed (discrete and continuous time) sequential model of the trading process, in which trades arise because of the interaction of three types of economic agents, *informed* and *uninformed traders* and a risk neutral, competitive *market maker*.

The magnitude of the bid-ask spread depends on the arrival rates of informed and uninformed traders which are governed by independent Poisson processes in their original framework and on the likelihood of the occurrence of three different types of information events ("no news", "good news" and "bad news") which are determined every day before the first trade takes place. Given the occurrence of a news event which has probability η_I, either a bad news event occurs with probability η_B or a good news event with probability $(1-\eta_B)$. 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 assumed to be determined by independent Poisson processes with identical arrival rate equal to λ_U. The probability structure of the trading process is summarized by the tree diagram in Figure 3.

However, if a news event occurs there will be additional order arrivals resulting from the transaction demand by informed traders, who are assumed to be risk neutral and competitive. Informed traders observe a signal, indicating either the presence of good or bad news, so their trade arrival rate will dependent upon the type of information event. 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. EKOP assume that two independent Poisson processes govern the arrival of informed buyers and sellers, both having the same arrival rate equal to λ_I.

Note, that the EKOP model implies, that trading evolves in different velocities, de-

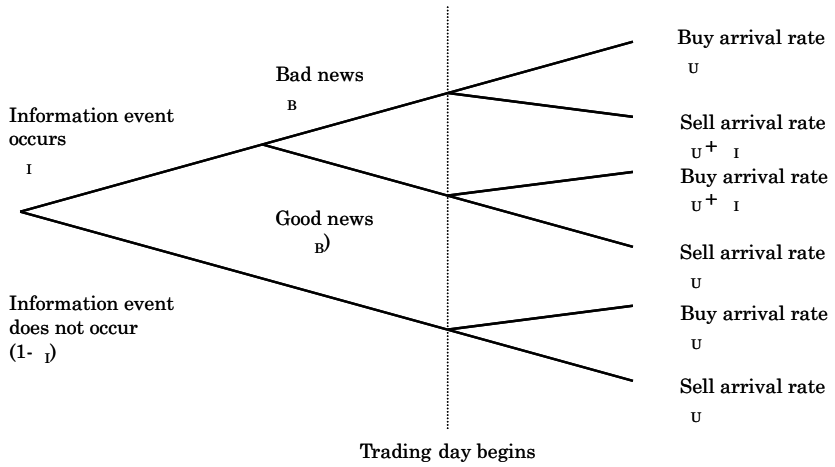

FIGURE 3: Structure of the trading process according to Easley, Kiefer, O'Hara and Paperman (1996).

pending on the type of the signal that has been observed by informed traders. It also implies that the data generating process of the trade durations will be a mixture of 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 motivated by a generalization of the EKOP model, in which it is assumed that (a) the information regime is not independent in time, but evolves according to a Markov chain during the trading day, (b) the arrival rates of both, uninformed traders and informed traders are not restricted to be the same for buyers and sellers, and (c) the conditional densities of the trade durations given the regime are not independent exponentials but rather follow a LACD(1,1) model, with marginal Burr density.

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

$$\tilde{b}_n = \gamma + \phi \cdot \cos(h(t_n)) + \delta \cdot \sin(h(t_n)) + \sum_{j=1}^{J-1} \beta_j \cdot \xi_{n|N}^{(j)} + \sum_{p=1}^{P} \varphi_p \cdot b_{n-p}, \tag{27}$$

where $\tilde{b}_n = p(b_n = 1)$ is the probability, that the n-th observed trade is buyer initiated, $\xi_{n|N}^{(j)}$ is the smoothed inference on the state of the regime variable s_n implied by the

estimated MSACD model presented in Section 4.3, b_n is the indicator variable, which is equal to one, if the n-th transaction was buyer initiated, and equal to zero, if it was seller initiated[22] and the sine and 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 $h(t)$ as defined in equation (26). 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.[23] This specification assumes, that the absolute value of the likelihood of being in regime j determines, whether the n-th trade will be more likely a buy or not.

We also consider a second specification of the regression function, in which the smoothed probabilities are replaced by log ratios

$$r_n^{(i,j)} = \ln\left(\frac{\xi_{n|N}^{(i)}}{\xi_{n|N}^{(j)}}\right), \qquad i \neq j$$

in the regression function. This specification stresses the importance of 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. If e.g. the probability of all three regimes is the same at some point in time, the log ratios will all be equal to zero, while if regime 1 has higher probability than regime 2 then the corresponding log ratio will be positive. If additionally the regression coefficient of $r_n^{(1,2)}$ is positive too, then the likelihood of

[22] We employed 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. Note that at the NYSE each trading day starts with a batch auction conducted by the delegated market maker, so an unknown number of trades after the open will result form these batch auctions, rather than from continuous trading. In our sample of Boeing 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.

[23] 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.

observing a buy will increase, whenever $\xi^{(1)}_{n|N}$ is greater than $\xi^{(2)}_{n|N}$.

Note that by comparing the magnitude and the sign of the two β coefficients we are always able to identify the nature of the information regime unambiguously in either of the two specifications. If, as in the preceding example, the log ratio of regime 1 and 2 has a positive coefficient, and 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 a number of different model specifications. The results of the estimation are presented in Table V.

Models 1, 2, and 3 include all possible combinations of the smoothed regime probabilities as regressors, models 4, 5, and 6 additionally include sine and cosine terms to control for time of day effects, and model 7 includes lags of b_n in the regression function.[24] In models 8, 9, and 10 the regime probabilities $\xi^{(j)}_{n|N}$ are replaced by the log ratios $r^{(i,j)}_n$. Note that the magnitudes and signs of the estimates of the β parameters in the first six models imply that the first regime is associated with good news, the second is bad news and the third is the no news regime, but with the exception of models 3 and 6 (where $\xi^{(2)}_{n|N}$ and $\xi^{(3)}_{n|N}$ appear as regressors) we find at least one of the parameter estimates to be insignificantly different from zero, as indicated by their t-statistic. Furthermore, when it is included, it is always the coefficient of the first regime $\xi^{(1)}_{n|N}$ that is found to be significant. We interpret this as evidence, that the model has no problems to classify the first regime. It seems to be clearly separated from the other two regimes. This is also consistent with the significance of $\xi^{(2)}_{n|N}$ and $\xi^{(3)}_{n|N}$ in models 3 and 6, as the corresponding coefficients only indicate differences from the reference category, which in the both models is associated with the omitted regressor, the probability of regime 1 $\xi^{(1)}_{n|N}$.

We therefore focus on the implied classifications of regimes 2 and 3 in models 7 to 10. Note that when lags of b_n are included as regressors, as in model 7, the implied

[24] 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.

classification of the regimes 2 and 3 changes, while the parameter estimates still indicate that regime 1 is the good news regime. Also, the parameters of the sine and cosine terms become insignificant. Note furthermore, that at the 5% significance level, both regime probabilities in model 7 are not significantly different from zero. We interpret this as an indication, that the absolute values of the regime probabilities might not be as important as their relative magnitudes. This intuition is confirmed by the regression results for models 8, 9, and 10, where both of the log ratios are significantly different from zero. All three specifications also 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. Thus, we conclude with high confidence, that this classification is the right one, and, since the parameters of the included log ratios are significantly different from zero, we find convincing evidence in favor of the generalized EKOP model.

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 stationary distribution of the Markov chain. These *ergodic probabilities* π_j can be interpreted as long run forecasts of the regime probabilities $\xi^{(j)}_{N+r|N}$ for $r \to \infty$.[25] The ergodic probabilities implied by our estimated 3-regime logarithmic Burr MSACD(1,1) model can be calculated from the transition probabilities of the Markov chain and are equal to $\pi_1 = 0.2873$, $\pi_2 = 0.5445$, and $\pi_3 = 0.1682$. Thus, if we stick to our classification of the regimes based on the probit estimates, the probability of informed trading in the sample period is equal to $1 - \pi_2 = 0.4555$, 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.

[25] See Hamilton (1994), p. 684.

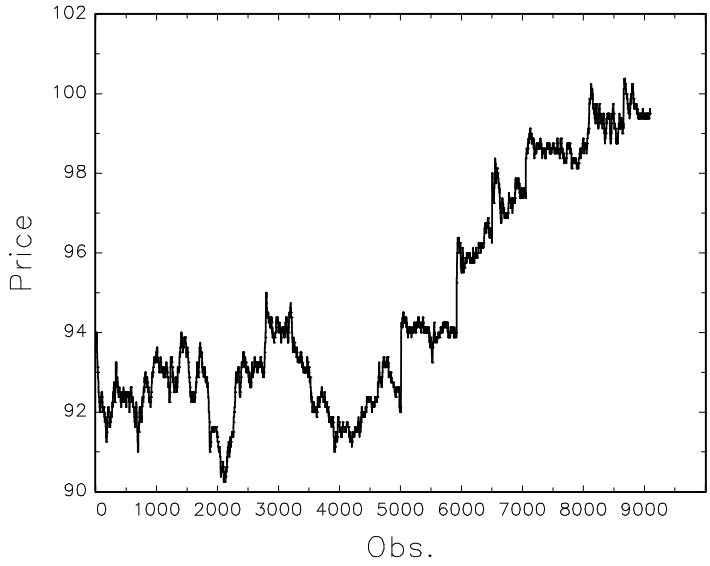

FIGURE 4: Evolution of the stock price of the common shares of Boeing during November 1996.

The price of the Boeing share rose from 93.625 US-$ at the beginning of our sample period to 99.500 US-$ at the end of November, see Figure 4. Most of the price changes are zero (69.20%), and there are approximately as many positive price changes (15.41%), as there are negative ones in our sample (15.29%), but the positive price changes appear slightly more concentrated in the second half of the sample, than in the first half and they are larger on average, when overnight returns are included.[26] Note that neither the evolution of the prices nor the price changes has been used to estimate any of the regime probabilities or to classify the resulting information regimes.

[26]The share of positive price changes that is larger than one tick (0.125 US-$) is equal to 0.36% and the maximum of the observed price changes is equal to 13 ticks (1.625 US-$), while the corresponding share for the negative price changes is only 0.30% and the maximum price drop is equal to -4 ticks (-0.500 US-$).

TABLE V: Estimation results for probit models

Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Constant	0.1111	2.2249	0.1079	4.3882	0.3578	7.1471	0.1117	2.1860	0.1087	4.3966
$\xi^{(1)}_{n\|N}$	0.2468	2.8847	0.2500	3.7778	-	-	0.2455	2.8549	0.2484	3.749
$\xi^{(2)}_{n\|N}$	-0.0032	-0.0612	-	-	-0.2500	-3.7778	-0.0030	-0.0587	-	-
$\xi^{(3)}_{n\|N}$	-	-	0.0032	0.0612	-0.2468	-2.8847	-	-	0.0030	0.0565
$r_n^{(2.1)}$	-	-	-	-	-	-	-	-	-	-
$r_n^{(2.3)}$	-	-	-	-	-	-	-	-	-	-
$\cos(h(t_n))$	-	-	-	-	-	-	-0.0313	-2.0041	-0.0313	-1.6402
$\sin(h(t_n))$	-	-	-	-	-	-	0.0401	2.1190	0.0401	2.0994
b_{n-1}	-	-	-	-	-	-	-	-	-	-
b_{n-2}	-	-	-	-	-	-	-	-	-	-
b_{n-3}	-	-	-	-	-	-	-	-	-	-
b_{n-4}	-	-	-	-	-	-	-	-	-	-
b_{n-5}	-	-	-	-	-	-	-	-	-	-
N	9067		9067		9067		9067		9067	
$\ln \mathcal{L}$	-6184.8		-6184.8		-6184.8		-6181.29		-6181.29	
$\ln \mathcal{L}_0$	-6191.9600		-6191.96		-6191.96		-6191.96		-6191.96	
LR_0	14.3209		14.3209		14.3209		21.3435		21.3435	
$P(LR_0)$	0.0008		0.0008		0.0008		0.0003		0.0003	
R^2_{MZ}	0.0025		0.0025		0.0025		0.0037		0.0037	
R^2_{AN}	0.0016		0.0016		0.0016		0.0023		0.0023	
R^2_{MF}	0.0012		0.0012		0.0012		0.0017		0.0017	
BIC	12396.94		12396.94		12396.94		12408.14		12408.14	

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^2_{MZ} is the value of the McKelvey and Zavoina R^2, R^2_{AN} is Aldrich and Nelson's R^2, and R^2_{MF} is McFadden's R^2. t-statistics have been computed based on QML estimates of the information matrix.

TABLE V: Estimation results for probit models (cont.)

Variable	Model 6		Model 7		Model 8		Model 9		Model 10	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Constant	0.3572	7.0751	-0.6549	-10.9844	0.1551	7.8313	0.1562	7.7974	-0.7724	-24.4609
$\xi_{n\|N}^{(1)}$	-	-	-	-	-	-	-	-	-	-
$\xi_{n\|N}^{(2)}$	-0.2484	-3.7490	-0.1190	-1.6594	-	-	-	-	-	-
$\xi_{n\|N}^{(3)}$	-0.2455	-2.8625	-0.1621	-1.7886	-	-	-	-		
$r_n^{(2.1)}$	-	-	-	-	-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
$\cos(h(t_n))$	-0.0313	-1.6402	-0.0112	-0.5454	-	-	-0.0322	-1.6852	-0.0117	-0.5685
$\sin(h(t_n))$	0.0401	2.0994	0.0204	0.994	-	-	0.0380	1.9913	0.0193	0.9396
b_{n-1}	-	-	0.9189	29.5350	-	-	-	-	0.9185	29.5184
b_{n-2}	-	-	0.3307	9.9744	-	-	-	-	0.3297	9.9441
b_{n-3}	-	-	0.2079	6.1627	-	-	-	-	0.2080	6.1672
b_{n-4}	-	-	0.1114	3.2775	-	-	-	-	0.1116	3.2802
b_{n-5}	-	-	0.0961	2.9807	-	-	-	-	0.0960	2.9754
N	9067		9062		9067		9067		9062	
$\ln \mathcal{L}$	-6181.29		-5170.79		-6180.93		-6177.57		-5169.55	
$\ln \mathcal{L}_0$	-6191.96		-6188.88		-6191.96		-6191.96		-6188.88	
LR_0	21.3435		2036.177		22.0605		28.7946		2038.657	
$P(LR_0)$	0.0003		0.0000		0.0000		0.0000		0.0000	
R^2_{MZ}	0.0037		0.2865		0.0039		0.005		0.2869	
R^2_{AN}	0.0023		0.1835		0.0024		0.0032		0.1837	
R^2_{MF}	0.0017		0.1645		0.0018		0.0023		0.1647	
BIC	12408.14		10432.70		12389.20		12400.70		10430.22	

6 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 inter 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 natural interpretation in the context of recent market microstructure models, suggesting that the existence of heterogenously informed investors is a major determinant of asset price variation. We showed how to use MSACD models for trade durations to test the implications of a generalized version of the market microstructure model by Easley, Kiefer, O'Hara, and Paperman (1996).

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.

A Appendix

A.1 Distributions for ACD models

TABLE VI: Distributional specifications for ACD models

	Exponential	Weibull	G. gamma[a,b]	Burr[a,c]
Parameters	$\left(\xi_n^{-1}\right)$	$\left(\xi_n^{-1},\gamma\right)$	$\left(\xi_n^{-1},\gamma,\nu\right)$	$\left(\xi_n^{-\kappa},\kappa,\sigma^2\right)$
Restrictions	$\xi_n>0$	$\xi_n,\gamma>0$	$\xi_n,\gamma,\nu>0$	$\xi_n,\kappa,\sigma^2>0$ and $\sigma^2<\kappa$
$\aleph$	1	$\frac{1}{\Gamma\left(1+\frac{1}{\gamma}\right)}$	$\frac{\Gamma(\nu)}{\Gamma\left(\nu+\frac{1}{\gamma}\right)}$	$\frac{\sigma^{2\left(1+\frac{1}{\kappa}\right)}\cdot\Gamma\left(\frac{1}{\sigma^2}+1\right)}{\Gamma\left(1+\frac{1}{\kappa}\right)\cdot\Gamma\left(\frac{1}{\sigma^2}-\frac{1}{\kappa}\right)}$
$f_n\left(x_n \mid \mathcal{F}_n;\theta\right)$	$\frac{1}{\xi_n}\cdot\exp\left(-\frac{x_n}{\xi_n}\right)$	$\frac{\gamma}{x_n}\cdot\left(\frac{x_n}{\xi_n}\right)^{\gamma}\cdot\exp\left(-\left(\frac{x_n}{\xi_n}\right)^{\gamma}\right)$	$\frac{\gamma}{\Gamma(\nu)\cdot x_n}\cdot\left(\frac{x_n}{\xi_n}\right)^{\gamma\cdot\nu}\cdot\exp\left(-\left(\frac{x_n}{\xi_n}\right)^{\gamma}\right)$	$\frac{\xi_n^{-\kappa}\cdot\kappa\cdot x_n^{\kappa-1}}{\left(1+\sigma^2\cdot\xi_n^{-\kappa}\cdot x_n^{\kappa}\right)^{\frac{1}{\sigma^2}+1}}$
$F_n\left(x_n \mid \mathcal{F}_n;\theta\right)$	$1-\exp(-\frac{x_n}{\xi_n})$	$1-\exp\left(-\left(\frac{x_n}{\xi_n}\right)^{\gamma}\right)$	$\frac{1}{\Gamma(\nu)}\Gamma\left(\nu,\left(\frac{x_n}{\xi_n}\right)^{\gamma}\right)$	$1-\left(1+\sigma^2\left(\frac{x_n}{\xi_n}\right)^{\kappa}\right)^{-\sigma^{-2}}$
$h_n(t \mid \mathcal{F}_n;\theta)^d$	$\frac{1}{\xi_n}$	$\frac{\gamma}{t-t_{n-1}}\cdot\left(\frac{t-t_{n-1}}{\xi_n}\right)^{\gamma}$	$\frac{\frac{\gamma}{t-t_{n-1}}\cdot\left(\frac{t-t_{n-1}}{\xi_n}\right)^{\gamma\cdot\nu}\cdot\exp\left(-\left(\frac{t-t_{n-1}}{\xi_n}\right)^{\gamma}\right)}{\Gamma(\nu)-\Gamma\left(\nu,\left(\frac{t-t_{n-1}}{\xi_n}\right)^{\gamma}\right)}$	$\frac{\xi_n^{-\kappa}\cdot\kappa\cdot(t-t_{n-1})^{\kappa-1}}{\left(1+\sigma^2\cdot\xi_n^{-\kappa}\cdot(t-t_{n-1})^{\kappa}\right)}$
$Var\left(x_n \mid \mathcal{F}_n;\theta\right)^e$	ψ_n^2	$\left(\aleph^2\cdot\Gamma\left(1+\frac{2}{\gamma}\right)-1\right)\cdot\psi_n^2$	$\left(\aleph^2\cdot\frac{\Gamma\left(\nu+\frac{2}{\gamma}\right)}{(\Gamma(\nu+))^2}-1\right)\cdot\psi_n^2$	$\left(\aleph^2\cdot\frac{\Gamma(1+\frac{2}{\kappa})\Gamma(\frac{1}{\sigma^2}-\frac{2}{\kappa})}{\sigma^{2(1+\frac{2}{\kappa})}\Gamma(\frac{1}{\sigma}+1)}-1\right)\cdot\psi_n^2$

[a] $\Gamma(y)=\int_0^{\infty}u^{y-1}e^{-u}du$ and $\Gamma(y,g)=\int_0^{g}u^{y-1}e^{-u}du$.

[b] When $\nu=1$ the generalized gamma reduces to the Weibull, while $\nu=1$ and $\gamma=1$ yields the exponential distribution.

[c] The Burr nests the Weibull as a limiting distribution when $\sigma^2\to 0$ and the exponential, when both $\sigma^2\to 0$ and $\kappa=1$.

[d] $h_n(t \mid \mathcal{F}_n;\theta)$ is the hazard function for $t_{n-1}\leq t\leq t_n$.

[e] For all distributions $E\left(x_n \mid \mathcal{F}_n;\theta\right)\equiv\psi_n=\frac{\xi_n}{\aleph}$.

A.2 The distribution of the residuals ε in the MSACD model

Starting with the marginal density of x_n given in (7), 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

$$g_n\left(\varepsilon_n \mid \mathcal{F}_n;\theta\right) = \psi_n \cdot \sum_{j=1}^{J} p(s_n = j \mid \mathcal{F}_n;\theta) \cdot f_x\left(\varepsilon_n \cdot \psi_n \mid s_n = j, \mathcal{F}_n;\theta\right), \tag{28}$$

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

$$E\left[\varepsilon_n \mid \mathcal{F}_n\right] = E\left[\frac{x_n}{\psi_n} \mid \mathcal{F}_n\right] = \frac{\psi_n}{\psi_n} = 1,$$

and thus independent of n as in a standard ACD model. Recall, that for a mixture density of the form $f\left(y\right) = \sum_{j=1}^{J} p(s=j) \cdot f_j\left(y|s=j\right)$ the raw (uncentered) moments μ_m' are given by[27]

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

In order to derive an expression for the variance of ε_n, we first define $Var(x_n \mid s_n = j, \mathcal{F}_n) \equiv \varrho_n^{(j)}$. 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 $E\left(x_n^2 \mid s_n = j, \mathcal{F}_n\right) = \varrho_n^{(j)} + \left(\psi_n^{(j)}\right)^2$ and so the regime independent second moment is $E\left(x_n^2 \mid \mathcal{F}_n\right) = \sum_{j=1}^{J} \xi_{n|n-1}^{(j)} \cdot \left(\varrho_n^{(j)} + \left(\psi_n^{(j)}\right)^2\right)$. Thus the regime independent variance of x_n is

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

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

$$\begin{aligned} Var(\varepsilon_n \mid \mathcal{F}_n) &= \frac{1}{\psi_n^2} \cdot Var(x_n \mid \mathcal{F}_n) \\ &= \sum_{j=1}^{J} \xi_{n|n-1}^{(j)} \cdot \frac{\varrho_n^{(j)}}{\psi_n^2} + \sum_{j=1}^{J} \xi_{n|n-1}^{(j)} \cdot \left(\frac{\psi_n^{(j)}}{\psi_n}\right)^2 - 1 \\ &= \frac{1}{\psi_n^2} \cdot E(x_n^2 \mid \mathcal{F}_n) - 1. \end{aligned} \tag{29}$$

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 (29) a sufficient condition for time invariance of $Var(\varepsilon_n \mid \mathcal{F}_n)$ 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.

[27] See Cameron and Trivedi (1998), p. 130.

REFERENCES

Allenby, G. M., R. P. Leone, and L. Jen (1999): "A dynamic model of purchase timing with application to direct marketing," *Journal of the American Statistical Association*, 94(446), 365–374.

Andersen, T. G., and T. Bollerslev (1997): "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4(2-3), 115–158.

Andrews, D. W. K. (1991a): "Asymptotic normality of series estimators for nonparametric and semiparametric regression models," *Econometrica*, 59(2), 307–345.

——— (1991b): "Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroscedastic errors," *Journal of Econometrics*, 47(2-3), 359–377.

Bauwens, L., and P. Giot (1997): "The logarithmic ACD model: An application to market microstructure and NASDAQ," Discussion paper, Université Catholique de Louvain.

Bauwens, L., and P. Giot (2000): "The moments of first-order Log-ACD models," Discussion paper, Université Catholique de Louvain.

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

Böhning, D., E. Dietz, R. Schaub, P. Schlattman, and B. G. Lindsay (1994): "The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family," *Annals of the Institute of Statistical Mathematics*, 46(2), 373–388.

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

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

Cameron, A. C., and P. K. Trivedi (1998): *Regression analysis of count data*. Cambridge University Press, Cambridge.

Cecchetti, S. G., P.-S. Lam, and N. C. Mark (1990): "Mean revision in equilibrium asset prices," *American Economic Review*, 80(3), 398–418.

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 (with discussion)," *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

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.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach (1997): "Regime switching with time-varying transition probabilities," in *Nonstationary Time Series Analysis and Cointegration*, ed. by C. P. Hargreaves, pp. 284–302. Oxford University Press, Oxford.

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.

Eastwood, B. J. (1991): "Asymptotic normality and consistency of seminonparametric regression estimators using an upwards F test truncation rule," *Journal of Econometrics*, 20(1), 151–181.

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 heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogenity," *Econometrica*, 51(2), 277–304.

Engle, R. F., and J. R. Russell (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.

——— (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-trigonometric regression," *Biometrika*, 77(1), 1–9.

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

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(1), 133–163.

Gallant, A. R. (1981): "On the bias in flexible functional forms and an essentially unbiased form," *Journal of Econometrics*, 20(2), 285–323.

Goldfeldt, S. M., and R. E. Quandt (1965): "Some tests for homocscedasticity," *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(1), 16–38.

Gray, S. (1996): "Modelling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42(1), 27–62.

GRITZ, R. M. (1993): "The impact of training on the frequency and duration of employment," *Journal of Econometrics*, 57(1), 21–51.

HAMILTON, J., AND R. SUSMEL (1994): "Autoregressive conditional heteroscedasticity and changes in regime," *Journal of Econometrics*, 64(1 & 2), 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(1), 39–70.

——— (1994): *Time series analysis*. Princeton University Press, Princeton.

——— (1996): "Specification testing in Markov-switching time-series models," *Journal of Econometrics*, 70(1), 127–157.

HECKMAN, J. J., AND B. SINGER (1982): "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52(2), 271–320.

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, Cambridge.

LEE, C. M. C., AND M. J. READY (1991): "Inferring trade direction from intraday data," *The Journal of Finance*, 46(2), 733–746.

LUNDE, A. (1999): "A generalized gamma autoregressive conditional duration model," Discussion paper, Aalborg University.

MACDONALD, I. L., AND W. ZUCCHINI (1997): *Hidden markov and other models for discrete-valued time series*. Chapman & Hall, London.

MADHAVAN, A. (2000): "Market microstructure: A survey," *Journal of Financial Markets*, 3(3), 205–258.

MUNDACA, B. G. (2000): "The effect of interventions on realignment," *Journal of International Financial Markets, Institutions & Money*, 10(3-4), 323–347.

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," Discussion paper, Graduate School of Business, University of Chicago.

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.

Rydén, T., T. Teräsvirta, and S. Åsbrink (1998): "Stylized facts of daily return series and the hidden Markov model," *Journal of Applied Econometrics*, 13(3), 217–244.

Schwarz, G. (1978): "Estimating the dimension of a model," *Annals of Statistics*, 6(2), 461–464.

Trivedi, P. K., and P. Deb (1997): "Demand for medical care by the elderly: A finite mixture approach," *Journal of Applied Econometrics*, 12(3), 313–336.

Turner, C. M., R. Startz, and C. R. Nelson (1989): "A Markov model of heteroscedasticity, risk, and learning in the stock market," *Journal of Financial Economics*, 25(1), 3–22.

Wang, P., I. M. Cockburn, and M. L. Puterman (1998): "Analysis of patent data - A mixed-Poisson-regression approach," *Journal of Business & Economic Statistics*, 16(1), 27–41.

Wang, P., M. L. Puterman, I. M. Cockburn, and N. Le (1996): "Mixed Poisson regression models with covariate dependent rates," *Biometrics*, 52, 381–400.

Wedel, M., W. S. Desarbo, J. R. Bult, and V. Ramaswamy (1993): "A latent class Poisson regression model for heterogenous count data," *Journal of Applied Econometrics*, 8(4), 397–411.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50(1), 1–25.

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.