

The use of the comprehensive family of distributions for the regime switching ACD framework

Reinhard Hujer ^{a,*}, Sandra Vuletić ^b

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

^b *University of Frankfurt/M.*

Version: 22 December 2004

Abstract

In recent methodological work the well known ACD approach, originally introduced by Engle and Russell (1998), has been supplemented by the involvement of an unobservable stochastic process which accompanies the underlying process of durations via a discrete mixture of distributions. The *Mixture* ACD model, emanating from the specialized proposal of De Luca and Gallo (2004), has proved to be a moderate tool for description of financial duration data. The use of one and the same family of ordinary distributions has been common practice until now. Our contribution incites to use the rich parameterized comprehensive family of distributions which allows for interacting different distributional idiosyncrasies.

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

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

1 Introduction

Investigating the microstructure of financial markets has become very popular over the last twenty years. Theoretical assertions concerning the behavior of market participants in the presence of asymmetric information are discussed

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

in many contributions. In this respect Easley et al. (1996) deliver a prominent approach. Statistical methodology will be employed in order to check empirically the validity of the implications of market microstructure models. Since rich transaction data sets are available containing detailed information about the timing of trades, prices, volume and other relevant characteristics for a wide range of financial securities, it is possible to get to the bottom of financial markets. Theory and the application of a tailor - made statistical instrument are combined in the elaboration of Kokot (2004).

New econometric methods appear rapidly and they experience an extensive application in the branch of finance. The autoregressive conditional duration model (ACD) introduced by Engle and Russell (1998) is an auspicious approach which couples the spirit of time series models with econometric tools for the analysis of transition data. Ultra high frequency data, stemming from transaction data sets and having the characteristic of irregular spacing in time, are ideal actuality for the use of the innovative framework. The ACD model is perfectly suitable for the analysis of dynamics of arbitrary events associated with the trading process along time, and the durations between successive occurrences of interesting market events are object of investigation.

As demonstrated by Bauwens et al. (2004) the periods of time elapsing between successive trades exhibit an idiosyncrasy which could not even be captured by extensions of the original model. For the first time the flexible Markov switching ACD model developed by Hujer et al. (2002) is capable of higher forecast accuracy of the trading process itself, but it requires much effort and computing power in estimation. We intend to introduce an alternative model with a parsimonious parameterization, called the *Mixture* ACD model (MACD), which also attains to good performance. Integral part of the MACD model is a latent discrete valued regime variable whose involvement can be justified in the light of recent market microstructure models. The unobservable regime can be associated with the presence (or absence) of private information about an asset's value that is initially available exclusively to a subset of informed traders and only eventually disseminates through the mere process of trading to the broader public of all market participants.

The manageable MACD model bears a resemblance to the general switching autoregression model introduced by Hamilton (1989) and nests many of the existing autoregression duration models as special cases. There are several models that are closely related to our approach as well. Despite the affinity

to the duration model given by De Luca and Gallo (2004), the MACD model differs substantially in the distributional assumption. It has the discrete mixture in common with the threshold ACD model introduced by Zhang et al. (2001).

This paper is structured as follows: The MACD model will be introduced in Section 2. Techniques for its estimation will be discussed and a specification test applicable to MACD models will be presented too. Moreover we establish a relationship to market microstructure theory. In an empirical application in Section 3 we present estimation results employing a transaction data set for the common share of Boeing traded on the New York Stock Exchange. Finally, in Section 4 we summarize our main results.

2 The Mixture ACD model

2.1 The basic framework

Let $x_n = t_n - t_{n-1}$ be the duration between the $(n - 1)$ -th and the n -th market event with deterministic conditional mean function

$$\psi_n = E(x_n | \mathcal{F}_{n-1}; \theta_\psi), \quad (2.1)$$

where the information set \mathcal{F}_{n-1} consists of all preceding durations up to time t_{n-1} and θ_ψ is the corresponding set of parameters. The Mixture ACD model, also referred to as MACD in the following, is defined by some linear or log linear recursion of this conditional mean. The essential of the MACD model is that the duration process x_n is accompanied by an unobservable stochastic process s_n composed of a sequence of discrete valued random variables with finite support $\mathfrak{J} = \{j \mid 1 \leq j \leq J, J \in \mathbb{N}\}$. The latent process s_n has the task to represent the regime in which the duration process x_n prevails at time t_n . The innovation process

$$\varepsilon_n = \frac{x_n}{\psi_n}, \quad (2.2)$$

has a known discrete mixture distribution with an unconditional expectation equal to one and invariant higher moments across the N observations considered in the sample. The density of each innovation has the following general form of appearance

$$g(\varepsilon_n; \theta_\varepsilon, \theta_\pi) = \sum_{j=1}^J \pi^{(j)} g(\varepsilon_n | s_n = j; \theta_\varepsilon^{(j)}), \quad (2.3)$$

where each nonnegative weight $\pi^{(j)}$ represents the probability for prevailing in state j and $\theta_\varepsilon^{(j)}$ is the corresponding parameter vector characterizing the conditional density of the innovation process driven in the j -th regime. The exponential, the Weibull, the Burr (1942), and the generalized gamma distribution, all of them nested in the comprehensive family of distributions, may be used in order to specify the regime specific distributions of the innovation process.

The expected value of each innovation is constrained to be equal to one and at the same time this expected value turns out to be a discrete mixture of regime specific expectations. This implies the maintenance of the equality

$$1 = \sum_{j=1}^J \pi^{(j)} E(\varepsilon_n | s_n = j; \theta_\varepsilon^{(j)}) \quad (2.4)$$

which does not require that all the regime specific expectations are equal to one. By the change of variable technique the relevant density for statistical inference is the duration's marginal density

$$f(x_n | \mathcal{F}_{n-1}; \theta) = \sum_{j=1}^J \pi^{(j)} f(x_n | s_n = j; \theta^{(j)}) \quad (2.5)$$

which depends on the parameter vector θ arising from the conjunction of $\theta^{(j)} = (\theta_\varepsilon^{(j)}, \theta_\psi)'$ for all $j \leq J$ and $\theta_\pi = (\pi^{(1)}, \dots, \pi^{(J)})$.

2.2 Estimation of the Mixture ACD model

For discrete mixture models there are two practices by which maximum likelihood estimates of the parameter vector θ may be obtained. The direct numerical maximization of the log-likelihood function

$$\mathcal{L}(\theta) = \sum_{n=1}^N \ln [f(x_n | \mathcal{F}_{n-1}; \theta)] \quad (2.6)$$

under the linear constraint $\sum_{j=1}^J \pi^{(j)} = 1$ and additional restrictions for warranty of equation (2.4), nonnegativity, stationarity and eventually for distributional parameters is the standard approach. Unfortunately, log-likelihood functions of mixture models are characterized by the existence of multiple local maxima. In order to catch the global maximum, repetition of estimation

with different start values is strongly recommended. Since standard maximization algorithms often fail or produce nonsensical results, maximum likelihood estimates for discrete mixture models are often obtained by the use of the robust Expectation-Maximization (EM) algorithm introduced by Dempster et al. (1977).

2.3 Statistical inference

Diebold et al. (1998) propose a method to test the forecast performance of general dynamic models. The idea behind this specification test has been extensively used by Bauwens et al. (2004) to compare different types of ACD models. Denote by $\{f(x_n | \mathcal{F}_{n-1}; \hat{\theta})\}_{n=1}^N$ the sequence of density forecasts evaluated using the parameter vector estimate $\hat{\theta}$ from some parametric model and denote by $\{f(x_n | \mathcal{F}_{n-1}; \theta)\}_{n=1}^N$ the sequence of densities corresponding to the true but unobservable data generating process of x_n . As shown by Rosenblatt (1952), under the null hypothesis $H_0 : \{f(x_n | \mathcal{F}_{n-1}; \hat{\theta})\}_{n=1}^N = \{f(x_n | \mathcal{F}_{n-1}; \theta)\}_{n=1}^N$, the sequence of empirical integral transforms

$$\hat{\zeta}_n = \int_{-\infty}^{x_n} f_n(u | \mathcal{F}_{n-1}; \hat{\theta}) du \quad (2.7)$$

will be uniform i.i.d. on the unit interval. Any statistical test for uniformity in the sequence of integral transforms can be used to assess the forecast performance of the model under consideration. Consider partitioning the support of ζ into K equally spaced bins and denote the number of observations falling into the k -th bin by \mathcal{N}_k . The confrontation of theoretical frequencies $\varsigma_k = \frac{1}{K}$ with observed relative frequencies $\hat{\zeta}_k = \frac{\mathcal{N}_k}{N}$ constitutes the fundament of the statistic

$$RT_{\zeta} = -2 \cdot \sum_{k=1}^K \mathcal{N}_k \cdot \ln \left[\frac{\varsigma_k}{\hat{\zeta}_k} \right] \quad (2.8)$$

which has a χ^2 distribution with $(K - 1)$ degrees of freedom under the null hypothesis. Checks for quantiles being equal to the population counterpart implied by the standard uniform distribution can be conducted as well. Let \mathcal{N}_p be the number of empirical integral transforms being less or equal than p , then the statistic

$$Q_{\zeta_p} = \frac{\mathcal{N}_p - N \cdot p}{\sqrt{N \cdot p \cdot (1 - p)}} \quad (2.9)$$

follows approximately the standard normal distribution under the null hypothesis $H_0 : \zeta_p = p$. The independence feature may be checked by computing the Ljung and Box (1978) test for the sequence of empirical integral transforms. The statistical tests for i. i. d. uniformity may be supplemented by graphical tools. Departures from uniformity can easily be detected using a histogram plot or quantile-quantile plot based on the sequence of $\hat{\zeta}_n$, while the autocorrelogram for $\hat{\zeta}_n$ can be used in order to assess the independence property.

2.4 Link to market microstructure theory

The modern literature on the microstructure of financial markets is unfolding in the style of Easley et al. (1996). The intercommunity of this broad literature is the presence of diverse types of market participants. The initial position is that the market participants are differentiated by the level of information which they harness privately. Consequently the trading mechanism will be discussed under the aspect of asymmetric information. The market development can be explored against the background of the coexistence and interaction of two categories of traders: informed traders catch a signal indicating that an asset is either overpriced or underpriced while uninformed traders, also called liquidity traders or followers, do not notice anything. The informed trader's strategy consists of making purchases and sales of assets in the immediate aftermath of the recognition of favorable and unfavorable signals. The informed traders encroach upon the market development conjunctly and trigger heaped transactions as soon as they bushwhack relevant news. Uninformed traders are insensible in regard to the information processing and retain the habitual trading activity.

The collectivity of transactions, carried out either by the large attendance of uninformed traders or by sporadic emersions of informed traders can be seen as a realization of a point process and the corresponding probability law that governs the occurrence of trades can be specified by a duration statistic. The presence of different traders acting on the financial market makes the embedding of a conglomerate of trader specific characteristics into the ordinary ACD framework, introduced by Engle and Russell (1998), reasonable. Because a specific transaction does not reveal by which type of trader it has been induced, the introduction of an underlying unobservable mixing variable with

discrete distribution is reasonable.

This simple theoretical background is excellently reflected in the MACD framework. Thereby the regime variable is in the capacity of the mixing variable and the mixing parameters can be interpreted as fractions of the different trader types acting on the market. The level of discrepancy between trader specific peculiarities in trading behavior can be easily regulated by adapting the parameters inside of equation (2.4). The instantaneous transaction rates turn out to be different across the trader categories and this is what we want to achieve primarily.

3 Empirical application

3.1 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 from the trades and quotes database provided by the NYSE Inc. The sampling period spans 19 trading days from November 1 to November 27, 1996. We used all trades observed during the regular trading day (9:30 - 16:00). Trades recorded up to five minutes after the opening have been excluded from our analysis because they are suspected of being parts of the initial batch auction. The trading times have been recorded with a precision measured in seconds. Observations occurring within the same second have been aggregated to one trade. In the final data set we removed censored observations: durations from the last trade of the day until the close and durations from the open until the first trade of the day.

It is well known that the length of the durations varies in a deterministic manner during the trading day that resembles an inverted U-shaped pattern. Engle and Russell (1997) propose to decompose the duration series into a deterministic time of day function $\Phi(t_{n-1})$ and a stochastic component x_n , so that the raw durations are generated from $\tilde{x}_n = x_n \cdot \Phi(t_{n-1})$. In order to remove the deterministic component we apply the two step method proposed by Engle and Russell (1997) in which the time of day function is estimated separately from other model parameters.¹ Dividing each raw duration \tilde{x}_n in

¹ Simultaneous ML-estimation as in Engle and Russell (1998) and Veredas et al. (2002) is also feasible. Engle and Russell (1998) report that both procedures give similar results if sufficient data is available.

Table 1
Descriptive Statistics for trade durations

Statistic	Raw durations \tilde{x}_n	Adj. durations x_n
Mean	48.4877	1.0012
Standard deviation	62.0190	1.1949
Minimum	1.0000	0.0141
First Quartile	10.0000	0.2317
Median	27.0000	0.5872
Third Quartile	61.0000	1.2984
Maximum	894.0000	16.1672
N	9012	9012
Ljung Box statistic	3763.052	1382.507

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

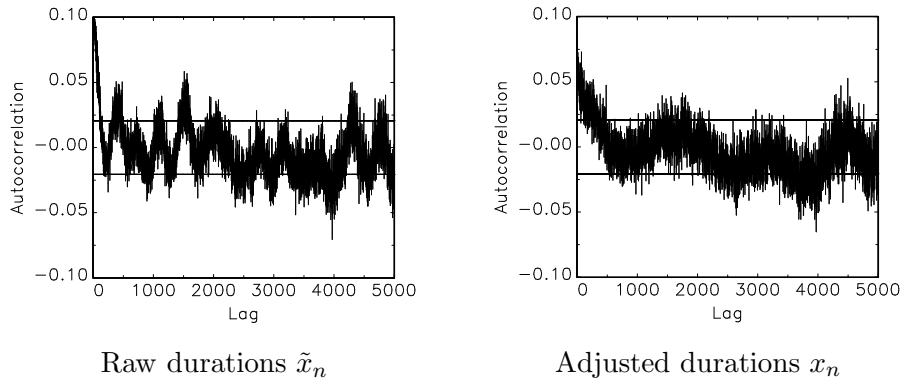
the sample by an estimate of the time of day function $\Phi(t_{n-1})$, a sequence of deseasonalized durations x_n is obtained which is used in all subsequent analyses.²

Descriptive information about sample moments and Ljung Box statistics of the raw and the seasonally adjusted duration data is reported in Table 1. As expected, the series of adjusted durations has a mean of approximately one. Both time series exhibit overdispersion relative to the exponential distribution which has standard error equal to mean. A mixture of distributions will accommodate well to the stylized fact of overdispersion.

Another eye-catching characteristic of the data is the presence of strong positive autocorrelation in the trade durations as can be seen in Figure 1. Even after seasonal adjustment, the Ljung-Box tests reject the hypothesis of no autocorrelation up to 50 lags at the 5% significance level, although the shape of the autocorrelation function changes slightly. Therefore, an autoregressive approach appears to be appropriate as a model for the transaction durations.

² Estimates of the time of day function were obtained by conducting a semi-nonparametric regression of the durations on the time of day according to Gallant (1981) and Eubank and Speckman (1990). Details on the seasonality adjustment step are available from the authors upon request.

Fig. 1. Autocorrelation function for durations



3.2 Model specification

The observed sequence of durations on a trading day will be treated independently of durations recorded on other trading days. This means that on every trading day a recursion determining the duration process starts anew. Consequently, the log likelihood function considering all available durations can be expressed as the sum of 19 daily log likelihoods. The mean function is chosen to be logarithmic and both lag orders p and q in the recursion are equal to one, i. e.

$$\psi_{d,n} = \exp(\omega) \cdot \psi_{d,n-1}^{\beta_1} \cdot x_{d,n-1}^{\alpha_1} \quad (3.1)$$

for $n \leq N_d$ and initial value $\psi_{d,1} = \bar{x}_d = \frac{1}{N_d} \sum_{n=1}^{N_d} x_{d,n}$ associated with each trading day $d \in \{1, \dots, 19\}$.

We estimate an ordinary ACD model and also a corresponding MACD model with consideration of two regimes. The ordinary ACD model is nested as a special case in the MACD framework with $J = 1$. Since the comprehensive family of distributions overcoats all customary duration distributions we zoom in on regime specific durations having density

$$f(x_{d,n} | s_{d,n} = j, \mathcal{F}_{d,n-1}; \theta^{(j)}) = \frac{[\nu_1^{(j)}]^{\frac{\nu_1^{(j)}}{2}} [\nu_2^{(j)}]^{\frac{\nu_2^{(j)}}{2}}}{B\left(\frac{\nu_1^{(j)}}{2}, \frac{\nu_2^{(j)}}{2}\right)} \cdot [\rho_{d,n}^{(j)}]^{\gamma^{(j)}} \gamma^{(j)} x_{d,n}^{\gamma^{(j)}-1} \cdot \frac{[\rho_{d,n}^{(j)} x_n]^{\gamma^{(j)} \left(\frac{\nu_1^{(j)}}{2} - 1\right)}}{[\nu_2^{(j)} + \nu_1^{(j)} \left(\rho_{d,n}^{(j)} x_{d,n}\right)^{\gamma^{(j)}}]^{\frac{\nu_1^{(j)} + \nu_2^{(j)}}{2}}} \quad (3.2)$$

with time-invariant degrees of freedom $\nu_1^{(j)}$ and $\nu_2^{(j)}$ entering the Beta function, regular time-invariant parameter $\gamma^{(j)}$ and time-variant parameter $\rho_{d,n}^{(j)} = \psi_{d,n}^{-1} \cdot \rho^{(j)}$. Both degrees of freedom are of major importance for characterizing the shape of the density and hazard rate. The Burr class of MACD models, introduced by Hujer and Vuletić (2004) by combining the distributional proposal of Grammig et al. (1998) and the mixture framework of De Luca and Gallo (2004), emerges by imposing the restriction $\nu_1^{(j)} = 2$ for every regime $j \leq J$. Thereby, the corresponding distributional parameters turn out to be $\mu_{d,n}^{(j)} = [\rho_{d,n}^{(j)}]^{\frac{1}{\gamma^{(j)}}}$, $\kappa^{(j)} = \gamma^{(j)}$ and $\sigma^{(j)} = 2 \cdot [\nu_2^{(j)}]^{-1}$. The distributional parameter $\kappa^{(j)}$ is the sole control lever of the hazard function shape for the j -th regime. For $\kappa^{(j)} \leq 1$ the Burr distribution implies a strong decreasing failure rate, while the case $\kappa^{(j)} > 1$ gives rise to a hunchbacked hazard function. Alternatively, when the second degree of freedom $\nu_2^{(j)}$ becomes very large then the density given in (3.2) describes approximately the generalized gamma distribution with parameters, $\lambda_{d,n}^{(j)} = \rho_{d,n}^{(j)} \cdot [0.5 \cdot \nu_1^{(j)}]^{\frac{1}{\gamma^{(j)}}}$, $\eta^{(j)} = \gamma^{(j)}$ and $\alpha^{(j)} = 0.5 \cdot \nu_1^{(j)}$. Different constellations for the parameters $\eta^{(j)}$ and $\alpha^{(j)}$ divide the shape property of the generalized gamma hazard function into the three general cases (constant, monotonic and nonmonotonic). The generalized gamma hazard rate is able to reproduce a decreasing (increasing) evolution in time as soon as the inequalities $\eta^{(j)} \cdot \alpha^{(j)} < 1$ and $\eta^{(j)} \leq 1$ ($\eta^{(j)} \cdot \alpha^{(j)} > 1$ and $\eta^{(j)} \geq 1$) hold true. Hunchbacked and bathtub graphs of the generalized gamma hazard function are also possible to obtain for $\eta^{(j)} \cdot \alpha^{(j)} > 1$, $\eta^{(j)} < 1$ and $\eta^{(j)} \cdot \alpha^{(j)} < 1$, $\eta^{(j)} > 1$ respectively. A constant hazard rate is obtained when the parameters satisfy the equalities $\eta^{(j)} \cdot \alpha^{(j)} = 1$ and $\eta^{(j)} = 1$ implying the exponential distribution as a special case. The use of the generalized gamma distribution for ACD modelling was initially advocated by Lunde (1999).

The regime specific distributions of a selective residual $\varepsilon_{d,n} = \psi_{d,n}^{-1} \cdot x_{d,n}$ are allowed to be nearly different. To be more concrete, all higher moments

$\mu_m^{(j)} = E\left(\varepsilon_{d,n}^m | s_{d,n} = j; \theta_\varepsilon^{(j)}\right)$ for arbitrary integer values $m > 1$ are generally regime specific but the fact $\mu_1^{(j)} = 1$ has to be in mind for every regime of interest. The following equalization

$$\rho^{(j)} = \frac{\Gamma\left(\frac{\nu_1^{(j)}}{2} + \frac{1}{\gamma^{(j)}}\right) \Gamma\left(\frac{\nu_2^{(j)}}{2} - \frac{1}{\gamma^{(j)}}\right)}{\Gamma\left(\frac{\nu_1^{(j)}}{2}\right) \Gamma\left(\frac{\nu_2^{(j)}}{2}\right)} \cdot \left[\frac{\nu_2^{(j)}}{\nu_1^{(j)}}\right]^{\frac{1}{\gamma^{(j)}}} \quad (3.3)$$

reflects the requirement of unit mean for every regime specific processes of innovations and ensures perennially the maintenance of condition (2.4) in the course of model estimation.

3.3 Estimation results

Parameter estimates and standard errors³ for all of the model specifications we estimated are presented in the upper panel of Table 2. With the aid of estimation results we carry out directly a couple of specification tests and we also calculate some informational measures. The values of test statistics and the corresponding p-values are given in the middle part of Table 2. The bottom of Table 2 comprehends values of the log-likelihood function and the Bayesian information criterion (*BIC*), which is computed as $-2 \cdot \mathcal{L} + \ln(N) \cdot k$ where k denotes the number of estimated parameters. We utilize some identifying notation in order to distinguish between different specifications which are appropriate candidates for framing a two-regime MACD model: the variable $D^{(j)}$ alludes to the distribution assumed for the j -th regime. The realization $D^{(j)} = \mathcal{C}$ indicates the use of the comprehensive distribution for the j -th regime, while the characters \mathcal{G} and \mathcal{B} stand for the generalized gamma distribution and the Burr distribution respectively.

³ Standard errors have been computed based on numerical derivatives of the incomplete log likelihood function using the quasi - maximum likelihood estimates of the information matrix as suggested by White (1982).

Table 2. Estimation results and specification tests for a one-regime and various two-regime MACD models

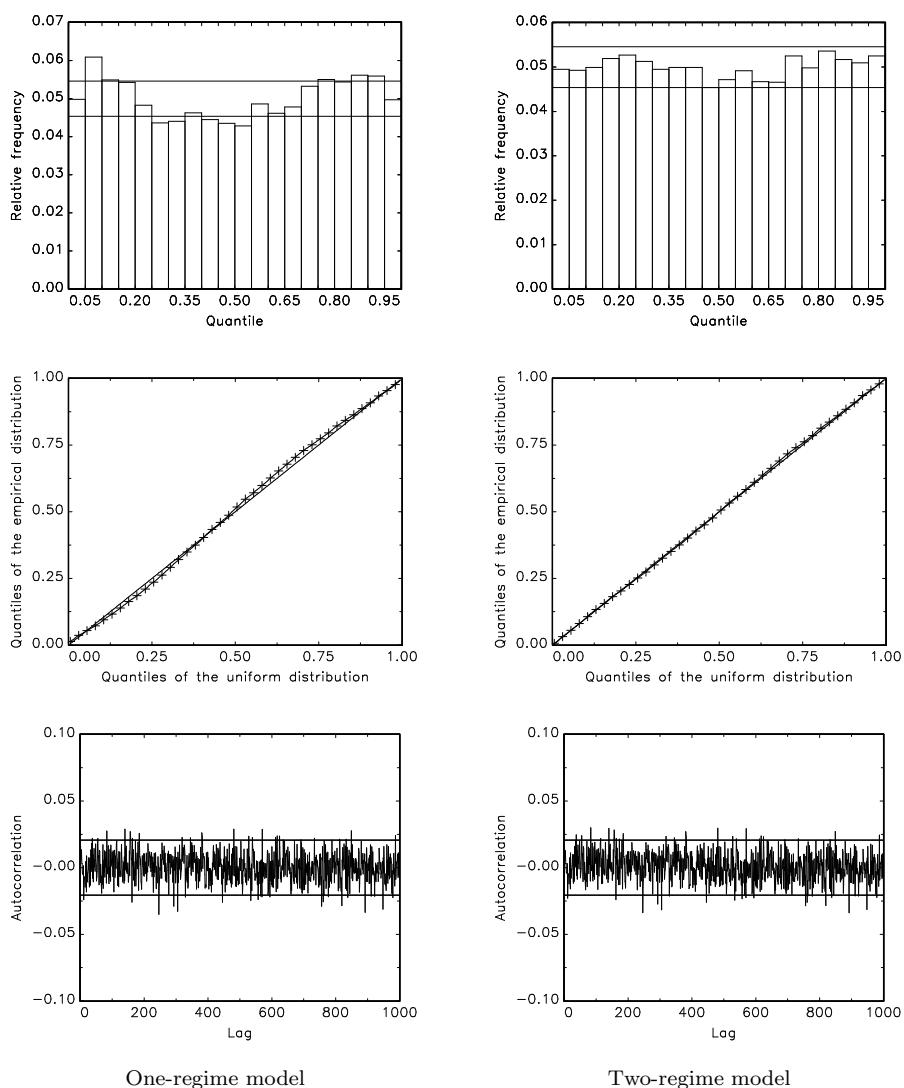
Parameter	$D^{(1)} = \mathcal{C}$		$D^{(1)} = \mathcal{C}, D^{(2)} = \mathcal{C}$		$D^{(1)} = \mathcal{G}, D^{(2)} = \mathcal{C}$		$D^{(1)} = \mathcal{C}, D^{(2)} = \mathcal{B}$		$D^{(1)} = \mathcal{G}, D^{(2)} = \mathcal{B}$	
	Estimate	Stderr	Estimate	Stderr	Estimate	Stderr	Estimate	Stderr	Estimate	Stderr
ω	0.022	0.003	0.031	0.004	0.031	0.004	0.031	0.004	0.031	0.004
α_1	0.038	0.004	0.041	0.005	0.041	0.005	0.041	0.005	0.041	0.005
β_1	0.949	0.008	0.940	0.010	0.940	0.010	0.939	0.010	0.939	0.010
$\eta^{(1)}$					0.435	0.035			0.412	0.035
$\gamma^{(1)}$	0.369	0.016	0.477	0.026			0.464	0.022		
$\kappa^{(2)}$							3.339	0.263	3.393	0.278
$\gamma^{(2)}$			2.024	0.446	1.997	0.413				
$\alpha^{(1)}$					5.337	0.774			5.906	0.890
$\nu_1^{(1)}$	12.593	1.042	9.338	0.877			9.822	0.799		
$\sigma^{(2)}$							3.100	0.258	3.154	0.273
$\nu_1^{(2)}$			5.657	3.187	5.989	3.315				
$\nu_2^{(1)}$	218.660	1.140	240.550	2.384			241.456	14.071		
$\nu_2^{(2)}$			1.077	0.244	1.091	0.232				
$\pi^{(1)}$			0.827	0.020	0.830	0.020	0.842	0.019	0.846	0.020
Statistic	Test	p-value	Test	p-value	Test	p-value	Test	p-value	Test	p-value
RT_ζ	94.606	0.000	17.281	0.571	17.061	0.586	21.538	0.308	21.061	0.333
LB_ζ	45.643	0.649	46.155	0.628	46.149	0.629	46.291	0.623	46.284	0.623
$Q_{\zeta 0.25}$	3.990	0.000	0.730	0.466	0.779	0.436	0.730	0.466	0.803	0.422
$Q_{\zeta 0.50}$	-1.875	0.061	-0.126	0.899	-0.169	0.866	-0.063	0.950	-0.169	0.866
$Q_{\zeta 0.75}$	-4.647	0.000	-1.873	0.061	-1.898	0.058	-1.946	0.052	-2.117	0.034
$\mathcal{L}(\hat{\theta}_{\{D^{(1)}, D^{(2)}\}})$		-8529.90		-8462.34		-8461.42		-8464.20		-8463.12
$BIC_{\{D^{(1)}, D^{(2)}\}}$		17114.44		17015.74		17004.79		17010.37		16999.10

We have two kind of investigations in mind. First of all, we are interested to examine the relation between the two-regime model specification that has conditional comprehensive distribution for durations in both regimes (labelled by $D^{(1)} = \mathcal{C}$, $D^{(2)} = \mathcal{C}$ in Table 2 and denoted by $\{\mathcal{C}, \mathcal{C}\}$ in the following discussion) and the corresponding one regime counterpart (labelled by $D^{(1)} = \mathcal{C}$). The incipient two-regime model specification $\{\mathcal{C}, \mathcal{C}\}$ will be reference when discussing other two-regime model specifications which are characterized by the feature of different distributional assumptions across the regimes.

Clearly, the *BIC* does not support the ordinary ACD model which is nested as a special case in the MACD framework. The test on the median argues for the null hypothesis $H_0 : \zeta_{0.5} = 0.5$, but the result is not in line with some strong conviction. The negligible p-values obtained from the other two quantile tests are sign of bad adaption in the tail of the distribution. Moreover, the alternative histogram specification test does not support the one regime model. This can be seen from the low p-value of the ratio test which is equal to zero. Hence, the apparent defect of the ordinary ACD model stems from the improper choice of distribution. However, the ordinary ACD model is able to capture the autocorrelation pattern of the intertrade durations adequately as indicated by the high p-value of the Ljung Box statistic for the series of empirical integral transforms. A significant improvement on the performance of the ordinary ACD model is obtained by allowing for interaction between a couple of regimes. Especially, the specification $\{\mathcal{C}, \mathcal{C}\}$ for the two-regime MACD model is able to eliminate the distributional problem of the ordinary ACD model and the autocorrelation pattern in the duration data will be still considered adequately. The p-value of the RT_ζ test and also the p-values of the first two quantile tests increase by leaps and bounds while $H_0 : \zeta_{0.75} = 0.75$ becomes statistical significant at the conventional significance level of 95 percent.

For purposes of comparison Figure 2 contains histogram plots, QQ-plots and graphs of the autocorrelation function for the series of integral transforms for the one regime model $\{\mathcal{C}\}$ and the two-regime model specification $\{\mathcal{C}, \mathcal{C}\}$. The plots clearly show that the estimated two-regime MACD model specification produces empirical integral transforms 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 the empirical integral transforms disagree sharply with the theoretical

Fig. 2. Histograms and QQ-plots for integral transforms



density, and that it tends to produce systematically biased forecasts for small and large durations. The histogram for a couple of quantiles is outside of the 95 percent confidence interval and a multitude of points are far from the diagonal in the QQ-plot. For both models, autocorrelations up to 1000 lags remain predominantly within the 95 percent confidence interval.

The primal two-regime model with comprehensive distribution of durations in both regimes is the easiest idea of multiple regime models which are in principle able to pass all the specification tests that we performed. The extraordinary improvement of the goodness of fit has been achieved by introducing four additional parameters compared to the one regime model. Therefrom, three parameters are required for the distributional matter while

the remaining parameter gets in touch with the regime probability. But possibly, the additional consideration of less than three distributional parameters makes the same fundamental result. In fact, improvement with no heavy losses is possible to reach by using a two-regime model specification that has two extra distributional parameters or even one (compare the results of specification tests given in the last three column blocks of Table 2). The usable reduction of distributional parameters reflects the use of the Burr or generalized gamma distribution instead of the comprehensive distribution, either for one of the two regimes or for both. The class of two-regime model specifications incorporating two extra distributional parameters (compared to the one regime model) is characterized by the feature that either the Burr or the generalized gamma distribution will be assumed for one regime while the assumption of comprehensive distributed durations retains for the other regime. Two-regime model specifications having only one extra distributional parameter result from using the Burr or generalized gamma distribution for both regimes.

As can be seen from the parameter estimates and standard errors, implied by the initial two-regime model specification $\{\mathcal{C}, \mathcal{C}\}$, the null hypothesis $H_0 : \nu_1^{(2)} = 2$ cannot be rejected even at the ten percent significance level. This makes an educated guess that the first degree of freedom in the second regime is equal to two. Consequently, the Burr density might be absolutely appropriate to describe the conditional distribution of durations in the second regime. The advantage of using the Burr distribution instead of the comprehensive distribution can be seen in the reduction of the number of distributional parameters. The estimation results of a MACD model having the comprehensive distribution in the first regime and the Burr distribution in the second regime, denoted by $\{\mathcal{C}, \mathcal{B}\}$ in the following, are gathered in the fourth block column of Table 2. The loss on likelihood when replacing the comprehensive distribution with the Burr distribution in the second regime is extremely small, i. e. the log-likelihood value $\mathcal{L}(\hat{\theta}_{\{\mathcal{C}, \mathcal{C}\}}) = -8462.32$ falls on the level $\mathcal{L}(\hat{\theta}_{\{\mathcal{C}, \mathcal{B}\}}) = -8461.42$ representing a relative change of 0.02 percent only. According to the *BIC*, the parsimonious model will be clearly preferred, because $17015.74 = BIC_{\{\mathcal{C}, \mathcal{C}\}} > BIC_{\{\mathcal{C}, \mathcal{B}\}} = 17010.37$.

Another eye-catching fact of the initial two-regime model specification $\{\mathcal{C}, \mathcal{C}\}$ is that the parameter estimate for the second degree of freedom in the first regime $\nu_2^{(1)}$ is extremely large. The estimation result $\hat{\nu}_2^{(1)} = 240.550$ and also the acceptance of the null hypothesis $H_0 : \nu_2^{(1)} \geq 200$ even at the ten per-

cent significance level justify the use of the generalized gamma distribution for the first regime. The third block column represents the estimation results we obtained for a MACD model with generalized gamma distribution for the first regime and comprehensive distribution for the second regime, denoted by $\{\mathcal{G}, \mathcal{C}\}$ in the following. This model specification is able to reduce the *BIC* as well, but the reduction is more heavy than in our first proposal of replacing the comprehensive distribution by the Burr distribution for the second regime. This spanking decrease comes into accordance with the increase of the value of the log-likelihood function with respect to the reference model. The increase of $\mathcal{L}(\hat{\theta}_{\{\mathcal{C}, \mathcal{C}\}}) = -8462.32$ by roughly 0.01 percent is plausible because the generalized gamma distribution results as a limiting case of the comprehensive distribution as soon as the second degree of freedom tends to infinity.

We feel impelled to combine the two partial proposals. So, we check out a two-regime model specification which is based on the assumption of generalized gamma distributed durations in the first regime and Burr distributed durations in the second regime, denoted by $\{\mathcal{G}, \mathcal{B}\}$ in the following. This specification is the most parsimonious one of all two-regime models we discussed until now and its estimation results are given in the last column block of Table 2. The *BIC* marks it as the best model. It seems as if the gain from extra complexity implied by the reference specification $\{\mathcal{C}, \mathcal{C}\}$ is small and the specification $\{\mathcal{G}, \mathcal{B}\}$ serves the purpose of better forecast performance effectively. Concerning the log-likelihood we find that the relation $\mathcal{L}(\hat{\theta}_{\{\mathcal{C}, \mathcal{B}\}}) < \mathcal{L}(\hat{\theta}_{\{\mathcal{G}, \mathcal{B}\}}) < \mathcal{L}(\hat{\theta}_{\{\mathcal{G}, \mathcal{C}\}})$ holds true, so that the two-regime model specification $\{\mathcal{G}, \mathcal{B}\}$ turns out to be a reasonable compromise between two-regime model specifications that assume the maintenance of the comprehensive distribution for one regime only.

The confrontation of each parsimonious specification with the reference specification $\{\mathcal{C}, \mathcal{C}\}$ will be carried out in order to obtain concrete information concerning the apportionment of gained (lost) likelihood by preferring the parsimonious specification. Let $\hat{f}_{\{D^{(1)}, D^{(2)}\}}^{(j)}(x_{d,n}) = f(x_{d,n} | s_{d,n} = j, \mathcal{F}_{d,n-1}; \hat{\theta}_{\{D^{(1)}, D^{(2)}\}}^{(j)})$ be the density characterizing the conditional distribution in the j -th regime of the two-regime model specification $\{D^{(1)}, D^{(2)}\}$ for $D^{(1)} = \mathcal{C}$ or $D^{(1)} = \mathcal{G}$ on the one hand and $D^{(2)} = \mathcal{C}$ or $D^{(2)} = \mathcal{B}$ on the other hand, and let $\hat{\pi}_{\{D^{(1)}, D^{(2)}\}}^{(j)}$ be the corresponding regime probability. Then we define for each regime the following set of functions

$$\hat{d}_1^{(j)}(x_{d,n}) = \hat{\pi}_{\{\mathcal{C}, \mathcal{B}\}}^{(j)} \cdot \hat{f}_{\{\mathcal{C}, \mathcal{B}\}}^{(j)}(x_{d,n}) - \hat{\pi}_{\{\mathcal{C}, \mathcal{C}\}}^{(j)} \cdot \hat{f}_{\{\mathcal{C}, \mathcal{C}\}}^{(j)}(x_{d,n}) \quad (3.4)$$

$$\hat{d}_2^{(j)}(x_{d,n}) = \hat{\pi}_{\{\mathcal{G}, \mathcal{C}\}}^{(j)} \cdot \hat{f}_{\{\mathcal{G}, \mathcal{C}\}}^{(j)}(x_{d,n}) - \hat{\pi}_{\{\mathcal{C}, \mathcal{C}\}}^{(j)} \cdot \hat{f}_{\{\mathcal{C}, \mathcal{C}\}}^{(j)}(x_{d,n}) \quad (3.5)$$

$$\hat{d}_3^{(j)}(x_{d,n}) = \hat{\pi}_{\{\mathcal{G}, \mathcal{B}\}}^{(j)} \cdot \hat{f}_{\{\mathcal{G}, \mathcal{B}\}}^{(j)}(x_{d,n}) - \hat{\pi}_{\{\mathcal{C}, \mathcal{C}\}}^{(j)} \cdot \hat{f}_{\{\mathcal{C}, \mathcal{C}\}}^{(j)}(x_{d,n}) \quad (3.6)$$

expressing the discrepancies between weighted regime specific likelihoods of competing two-regime model specifications discussed above. A visual impression on all these functions is given in Figure 3 which makes the graph of $\hat{d}_r^{(j)}(x_{d,n})$ available in its r -th row and j -th column. Note, that large durations are relative insensitive to an arbitrary change of the distributional assumption, while small durations tend to react heavily. Another distinctive feature seems to be that the amplitude of absolute likelihood changes for the first regime is lower than the corresponding amplitude for the second regime, but $\hat{d}_r^{(1)}(x_{d,n})$ needs more time to draw near zero. Because of the salient fact of stable probability estimates across all model specifications involving two regimes we can conclude that any parsimonious specification gives tendentially more likelihood to the first regime compared to the corresponding likelihood of the rich parameterized reference specification $\{\mathcal{C}, \mathcal{C}\}$. At the same time the second regime takes a loss concerning the likelihood. Consequently, we have two contrary effects acting on the change of the log likelihood value when passing from the reference specification $\{\mathcal{C}, \mathcal{C}\}$ into a parsimonious specification.

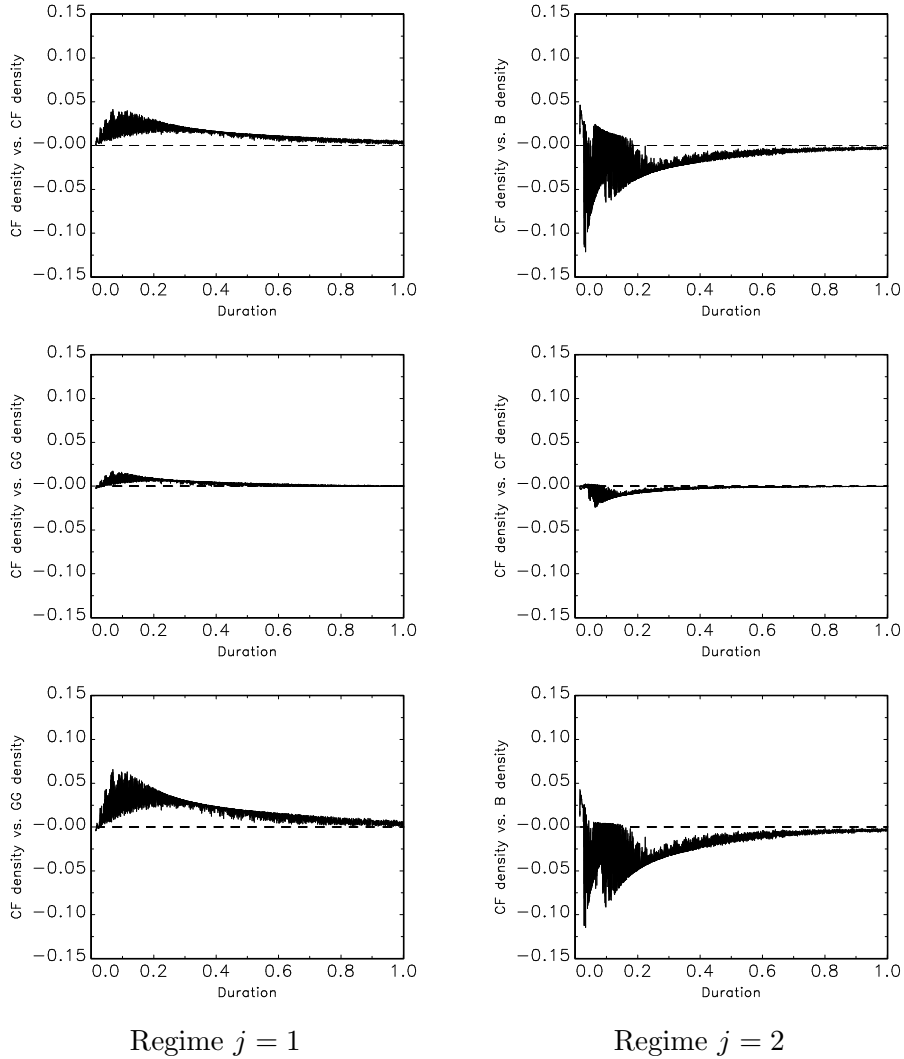
The dominance of one or the other effect depends on the choice for parsimonious specification and an elaborate discussion can be conducted by using the measures

$$\ell_r^{(j)} = \sum_{d=1}^{19} \sum_{n=1}^{N_d} \hat{d}_r^{(j)}(x_{d,n}) \quad (3.7)$$

$$h_r^{(j)}(w) = \frac{\sum_{d=1}^{19} \sum_{n=1}^{N_d} 1_{\{|\hat{d}_r^{(j)}(x_{d,n})| > w\}}}{\sum_{d=1}^{19} N_d} \cdot 100 \quad (3.8)$$

for $r \leq 3$ and $j \leq 2$ and appropriate non-negative values for w . A comparison between $\ell_r^{(1)}$ and $\ell_r^{(2)}$ with respect to the magnitude and sign and the discrepancy between $h_r^{(1)}(w)$ and $h_r^{(2)}(w)$ are conducive to find the scene from which some log-likelihood change runs out mainly. Table 3 collects all relevant measures, we decide to use $w = 0.05$ in order to catch upper outliers of $\hat{d}_r^{(j)}(x_{d,n})$, while the alternative $w = 0.01$ cares for non-extremal values. For the first specification adjustment we find that the inequality $\ell_1^{(1)} < -\ell_1^{(2)}$

Fig. 3. Likelihood differences between different two-regime model specifications



holds true which means that the replacement of the comprehensive distribution with the Burr distribution for the second regime is responsible for the log-likelihood loss registered previously. The value $h_1^{(2)}(0.05) = 2.142$ gives information that this log-likelihood loss is predominantly caused by a small number of durations coming along with wide differences between regime specific likelihood contributions. These durations are typically extremely small. The situation for the second specification transfer is different from the first. The fact $\ell_2^{(1)} > -\ell_2^{(2)}$ implies that the gained log-likelihood is caused by replacing the comprehensive distribution with the generalized gamma distribution for the first regime. The log-likelihood gain results from the majority of observations ($100 - h_2^{(1)}(0.01) = 95.628$ percent) having marginal differences between regime specific likelihood contributions. For the omnibus specification transfer

Table 3
Informative mesures

Measure	Regime $j = 1$	Regime $j = 2$
$\ell_1^{(j)}$	96.619	-98.120
$\ell_2^{(j)}$	26.145	-25.163
$\ell_3^{(j)}$	137.807	-136.525
$h_1^{(j)}(0.01)$	48.746	41.345
$h_2^{(j)}(0.01)$	4.372	8.855
$h_3^{(j)}(0.01)$	52.985	44.097
$\hat{h}_1^{(j)}(0.02)$	18.065	24.223
$\hat{h}_2^{(j)}(0.02)$	0.000	0.000
$\hat{h}_3^{(j)}(0.02)$	37.261	30.681
$h_1^{(j)}(0.05)$	0.000	2.142
$h_2^{(j)}(0.05)$	0.000	0.000
$h_3^{(j)}(0.05)$	1.198	5.870

we find $\ell_3^{(1)} > -\ell_3^{(2)}$ even though we observed a loss of the log-likelihood value. But this contradiction can be explained by the concave increase of the logarithm function. The function $\hat{d}_3^{(1)}(x_{d,n})$ has slower convergence to zero than $\hat{d}_3^{(2)}(x_{d,n})$. The fraction of values $|\hat{d}_3^{(1)}(x_{d,n})|$ being greater than 0.01 is equal to 52.985 percent, while the corresponding fraction amounts to 44.097 percent for the second regime.

The parameter estimates for ω , α_1 and β_1 , which determine the evolution of the duration's conditional mean in time, differ only marginally across the four two-regime model specifications we estimated. The same fact may be noticed for the distributional parameters. The estimation results obtained from the reference model specification $\{\mathcal{C}, \mathcal{C}\}$ show that the three regular distributional parameters $\gamma^{(j)}$, $\nu_1^{(j)}$ and $\nu_2^{(j)}$ vary keenly across the regimes. Both estimated degrees of freedom have larger values in the first regime than in the second and we find that $\hat{\gamma}^{(1)} < \hat{\gamma}^{(2)}$ holds true. This has a strong impact on the shape of the hazard function considered for each regime separately. The regime specific hazard function

$$\hat{\lambda}_{\{D^{(1)}, D^{(2)}\}}^{(j)}(x_{d,n}) = \frac{\hat{f}_{\{D^{(1)}, D^{(2)}\}}^{(j)}(x_{d,n})}{1 - \int_0^{x_{d,n}} \hat{f}_{\{D^{(1)}, D^{(2)}\}}^{(j)}(u) du} \quad (3.9)$$

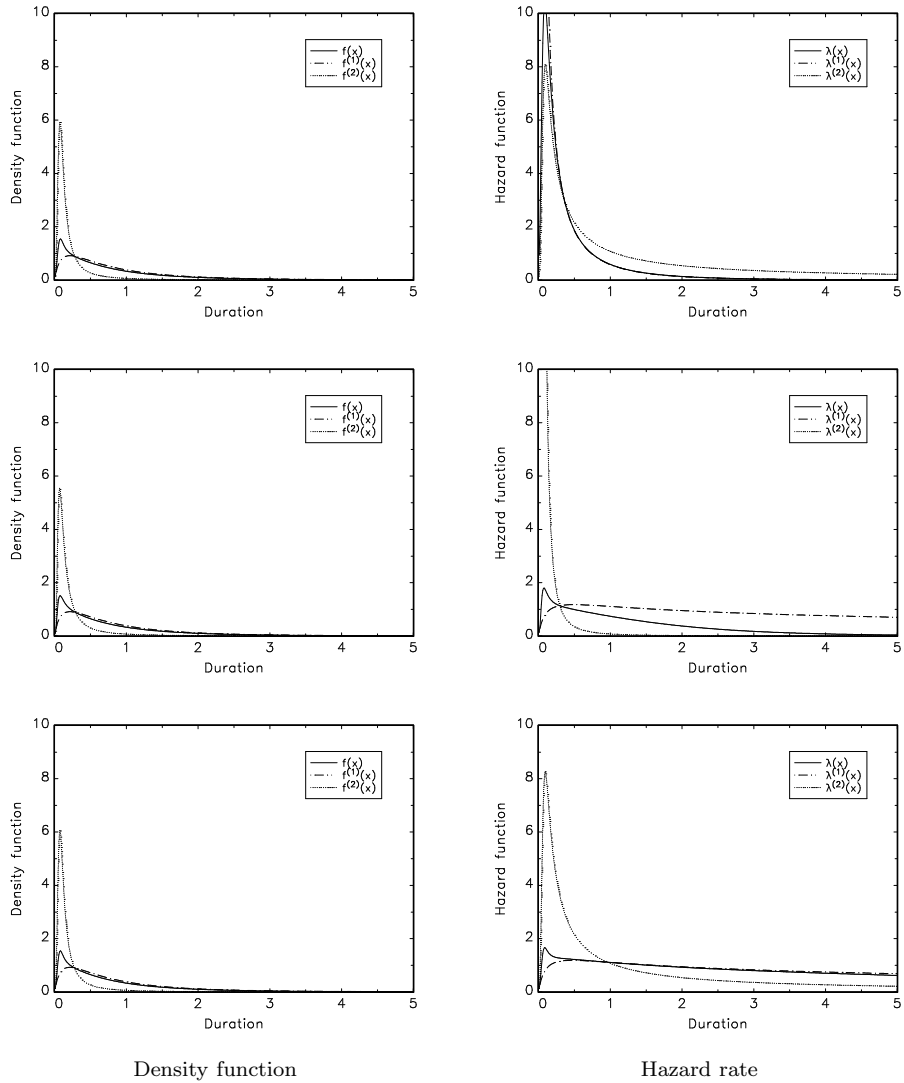
for each regime $j \leq 2$ and also the regime unspecific hazard rate

$$\hat{\lambda}_{\{D^{(1)}, D^{(2)}\}}(x_{d,n}) = \frac{\sum_{j=1}^J \hat{\pi}^{(j)} \cdot \hat{f}_{\{D^{(1)}, D^{(2)}\}}^{(j)}(x_{d,n})}{\sum_{j=1}^J \hat{\pi}^{(j)} \cdot \left[1 - \int_0^{x_{d,n}} \hat{f}_{\{D^{(1)}, D^{(2)}\}}^{(j)}(u) du \right]} \quad (3.10)$$

evaluated for $\psi_{d,n} = 1$ are displayed on the right hand side of Figure 4 representing the case of two-regime model specification $\{D^{(1)}, D^{(2)}\} = \{\mathcal{C}, \mathcal{B}\}$ ($\{D^{(1)}, D^{(2)}\} = \{\mathcal{G}, \mathcal{C}\}$) [$\{D^{(1)}, D^{(2)}\} = \{\mathcal{G}, \mathcal{B}\}$] in its first (second) [third] row, and the corresponding densities are given on the left hand side. Note in the first instance, that the choice for the one or the other specification does not change the qualitative nature of the density. We observe the maintenance of $\hat{\eta}^{(1)} \cdot \hat{\alpha}^{(1)} > 1$ and $\hat{\eta}^{(1)} < 1$ so that the hazard rate of the first regime turns out to be hunchbacked for generalized gamma distributed durations. In analogy, the hazard function characterizing the second regime is hunchbacked as well because of $\hat{\kappa}^{(2)} > 1$ for the Burr distribution. For each parsimonious two-regime model specification the hazard rate assigned to the second regime tends to rise rather quickly after a transaction has been observed. In contrast the hazard function under the first regime increases moderately and gives clearly more weight to larger spells for the specifications $\{\mathcal{G}, \mathcal{C}\}$ and $\{\mathcal{G}, \mathcal{B}\}$. This corresponds nicely to the fact that the first regime has higher probability $\hat{\pi}^{(1)}$ than the second regime. Roughly 80 percent of all transactions were generated in the first regime. The application of the MACD model affirms the existence of two constitutively different streams governing the process of intertrade durations and visualizes the different velocities from which trading evolves. The inertial trading activity, adumbrated by the hazard rate of the first regime, predominates the whole trading process and can be associated with the theoretical vision of trading behavior ascribed to the uninformed traders. The second regime awards the image of succinct trading which can be traced back to informed traders participating on the financial market.

Summarizing, our application illustrates that the conditional distribution of durations in the first regime is generalized gamma while durations in the second regime follow rather the Burr distribution. This empirical experience

Fig. 4. Density and hazard function



makes the usual strategy of using one common distribution family for all regimes problematic. Limitations concerning the intensity rate would be an unavoidable consequence. An attractive possibility to avoid problems coming from a distributional misspecification will be the use of the comprehensive family of distributions which allows for extraordinary flexibility.

4 Conclusions

Mixture models are frequently used in econometrics. So, researcher feel defiant to combine the methodological background of mixture models with the art of ACD modelling, originally introduced by Engle and Russell (1998).

Both, our discrete mixture ACD model which traces back to the basic concept of De Luca and Gallo (2004) and the Markov Switching ACD model of Hujer et al. (2002) act as a promising new approaches for modelling auto-correlated durations obtained from high frequency data sets from stock and foreign exchange markets. They are able to remove the distributional problem from which ordinary ACD models occasionally suffer. A further asset of these models is that they can be interpreted in the context of recent market microstructure models.

But until now one and the same family of distributions has been assumed for specifying all regime specific densities of durations within the framework of regime switching ACD models. Typically, either the class of Burr distributions or the class of generalized gamma distributions has come into consideration so far, as done by Hujer and Vuletić (2004) and Liu et al. (2004). The idea of using an all-embracing distribution, which nests common waiting time distributions as special cases, is the innovation we would like to provide. A distribution belonging to the comprehensive family is rich in parameters but allows for best customization. Moreover it makes possible to detect secret distributional nature for each regime of interest.

References

- Bauwens, L., Giot, P., Grammig, J. Veredas, D., 2004. A comparison of financial duration models via density forecasts. *International Journal of Forecasting* 20, 589–609, forthcoming.
- Burr, I. W., 1942. Cumulative frequency functions. *Annals of Mathematical Statistics* 13, 215–232.
- De Luca, G., Gallo, Giampiero, M., 2004. Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics & Econometrics* 8 (2), 1–18, article 8.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39, 1–38, series B.
- Diebold, F. X., Gunther, T. A., Tay, A. S., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39 (4), 863–883.
- Easley, D., Kiefer, N., O’Hara, M., Paperman, J. P., 1996. Liquidity, informa-

- tion and infrequently traded stocks. *Journal of Finance* 51 (4), 1405–1436.
- Engle, R. F., Russell, J. R., 1997. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of Empirical Finance* 4 (2-3), 187–212.
- Engle, R. F., Russell, J. R., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66 (5), 1127–1162.
- Eubank, R. L., Speckman, P., 1990. Curve fitting by polynomial-trigonometric regression. *Biometrika* 77 (1), 1–9.
- Gallant, A. R., 1981. On the bias in flexible functional forms and an essentially unbiased form. *Journal of Econometrics* 20 (2), 285–323.
- Grammig, J., Hujer, R., Kokot, S., Maurer, K. O., 1998. Modeling the deutsche telekom ipo using a new acd specification - an application of the burr-acd model using high frequency ibis data. Discussion Paper 55, Sonderforschungsbereich 373, Humboldt Universitt zu Berlin.
- Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57 (2), 357–384.
- Hujer, R., Vuletić, S., 2004. Econometric analysis of financial trade processes by mixture duration models, j. W. Goethe University, Frankfurt on the Main.
- Hujer, R., Vuletić, S., Kokot, S., 2002. The markov switching ACD model. Working Paper Series: Finance and Accounting 90, University of Frankfurt.
- Kokot, S., 2004. *The Econometrics of Sequential Trade Models : Theory and Applications Using High Frequency Data*. Springer Verlag.
- Liu, Y., Hong, Y., Wang, S., 2004. How well can autoregressive duration models capture the price durations dynamics of foreign exchanges. Tech. Rep. 2004007 EN, The China Center for Financial Research, Tsinghua University.
- Ljung, G. M., Box, G. E. P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65 (2), 297–303.
- Lunde, A., 1999. A generalized gamma autoregressive conditional duration model, Department of Economics, Politics and Public Administration, Aalborg University, Denmark.
- Rosenblatt, M., 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23 (3), 470–472.
- Veredas, D., Rodriguez-Poo, J., Espasa, A., 2002. On the intradaily seasonal-

- ity and dynamics of a financial point process: A semiparametric approach. Discussion Paper 23, CORE, Université Catholique de Louvain.
- White, H., 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1), 1–25.
- Zhang, M. Y., Russell, J. R., Tsay, R. S., 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* 104 (1), 179–207.