Econometrics Journal (2000), volume **3**, pp. 16–38.

Non-monotonic hazard functions and the autoregressive conditional duration model

JOACHIM GRAMMIG, KAI-OLIVER MAURER

Department of Economics and Business Administration, University of Frankfurt
E-mail: grammig@wiwi.uni-frankfurt.de; K.Maurer@em.uni-frankfurt.de

Received: August 1999

Summary This paper shows that the monotonicity of the conditional hazard in traditional ACD models is both econometrically important and empirically invalid. To counter this problem we introduce a more flexible parametric model which is easy to fit and performs well both in simulation studies and in practice. In an empirical application to NYSE price duration processes, we show that non-monotonic conditional hazard functions are indicated for all stocks. Recently proposed specification tests for financial duration models clearly reject the standard ACD models, whereas the results for the new model are quite favorable.

Keywords: *Financial transactions data, Autoregressive conditional duration model, Hazard function, Burr distribution, Market microstructure, Price durations, Self-exciting point process.*

1. INTRODUCTION

The market microstructure papers by Easley *et al.* (1996), Diamond and Verrecchia (1987), Glosten and Milgrom (1985), Hasbrouck (1988) and O'Hara (1995) emphasize that the waiting times between events such as trades, quote updates, price changes, and order arrivals play a key role in understanding the processing of private and public information in financial markets. Hence, the accessibility of financial transactions data, i.e. real time recordings of trades, order arrivals and quote updates, opened new perspectives for the empirical analysis of market microstructure processes. By appropriately editing the data it is possible to define almost any event of interest, and the corresponding duration sequence.

An econometric framework for the modeling of financial duration processes with intertemporally correlated event arrival times has been provided by Engle and Russell (1998) who have introduced the Autoregressive Conditional Duration (ACD) model. The ACD model belongs to the family of self-exciting marked point processes originally proposed by Cox and Lewis (1966), Hawkes, (1971a, 1971b, 1972) and Rubin (1972). A point process is a sequence of event arrival times $\{t_0, t_1, \ldots, t_n, \ldots\}$ with $t_0 < t_1 < \cdots < t_n \cdots$, and an associated function $N(t)$ counting the number of events that have occurred by time t. A point process is described as self-exciting when the the past evolution impacts the probability of future events. The marks contain the information that is associated with these events. Perceiving a sequence of trades in an intra-day financial market as a marked point process, the marks include events such as transaction prices and the traded volumes.

By introducing the ACD approach Engle and Russell (1998) revitalized the interest in 'time series models of time' that had developed after the early papers by Wold (1948) and Cox (1955), and the applications of Hawkes's ideas to earthquake data by Ogata and Katsura (1986), Vere-Jones and Ozaki (1982), and the introduction of ARMA-type models by Jacobs and Lewis (1977), Lawrence and Lewis (1980) and Gaver and Lewis (1980). The ACD approach adopts elements from conditional heteroskedasticity modeling, and the reader familiar with the papers by Engle (1982) and Bollerslev (1986) will recognize many parallels. Basically, the economic motivation behind the ACD and the ARCH model follows a similar logic: due to a clustering of news, financial market events occur in clusters. This implies that the waiting times between these events exhibit a significant serial correlation.

Recently, several extensions to Engle and Russell's standard models have been proposed. Engle (2000) and Ghysels and Jasiak (1998a) combine conditional duration models with GARCH-type effects (ACD-GARCH), whereas Ghysels *et al.* (1998) introduce a stochastic volatility duration model to cope with higher order dynamics in financial duration processes. Ghysels and Jasiak (1998b) investigate the persistence of inter-trade durations using a fractionally integrated ACD model, whilst Zhang *et al.* (1999) advocate a non-linear version of the ACD model rooted in a self-exciting threshold autoregressive framework. The logarithmic ACD model introduced by Bauwens and Giot (1997) provides a robust framework for testing market microstructure hypotheses as it avoids parameter restrictions implied by the original ACD specification. Gerhard and Hautsch (1999) address the problems associated with grouped financial duration data. Bauwens and Giot (1998) and Russell and Engle (1998) propose extensions to deal with competing risks, whereas Russell (1998) and Engle and Lunde (1998) consider bivariate models for trade and quote processes. An comparison of financial duration models based on density forecast evaluation methods is conducted by Bauwens *et al.* (2000).

An idiosyncrasy of Engle and Russell's standard ACD models is that the implied hazard (or intensity) functions conditional on past durations are restricted to being either constant, increasing or decreasing with respect to duration. Independently from our work Bauwens and Veredas (1999), Lunde (1999), Hamilton and Jorda (1999) and Zhang *et al.* (1999) have questioned whether imposing these restrictions is appropriate, and have proposed specifications that offer greater flexibility. In this paper we introduce an alternative ACD specification and address the question: what consequences might a misspecification of the hazard function have? We are especially concerned with investigating whether the expected duration forecasts are affected. This is of key importance for the class of ACD-GARCH models, in which expected inter-trade durations enter the conditional heteroskedasticity equation as pre-determined variables.

In a Monte Carlo study we show that the misspecification of the hazard function can severely deteriorate the ACD model's ability to predict expected durations. An empirical application employs alternative ACD models for an analysis of price duration processes at the New York Stock Exchange (NYSE). For three reasons quote durations can be considered as the most interesting financial duration process: first, there is a link with the instantaneous volatility of the quoted mid-price process, as pointed out by Engle and Russell (1998); second, as shown by Pringent *et al.* (1999), the behavior of price durations has important implications for option pricing; third, the price duration process can be used to empirically test microstructure theories as demonstrated by Bauwens and Giot (1998) and Engle and Russell (1998). We find that for all stocks considered, non-monotonic hazard functions are indicated. Applying recently proposed specification tests for financial duration models, the standard ACD models are clearly rejected, whereas the results for the new specification are quite encouraging.

The remainder of the paper is organized as follows: Section 2 provides the motivation to deal

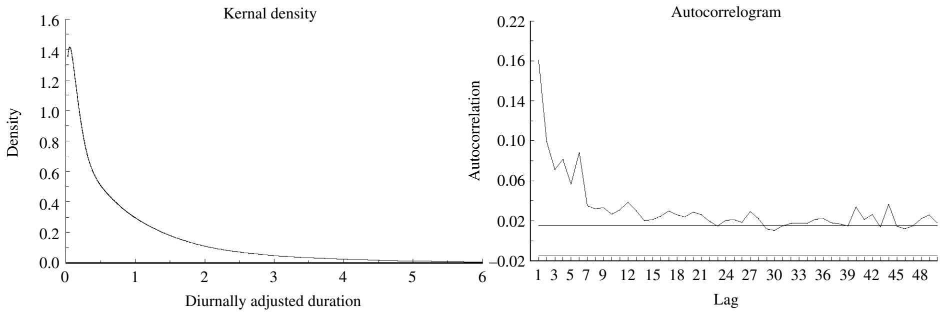

Figure 1. Density estimate using a Gamma kernel and autocorrelogram for diurnally adjusted IBM trade durations at NYSE November 1, 1990–November 30, 1990. As suggested by Chen (1999) the bandwidth was chosen as $h = \{(0.9sN)^{-0.2}\}^2$, where s denotes the standard deviation and N the length of the duration sequence.

with non-monotonic hazard functions when modeling financial duration processes; Section 3 introduces the new specification after a brief review of the ACD approach; Section 4 presents the results of the Monte Carlo study; Section 5 contains the empirical applications; Section 6 concludes.

2. MOTIVATION

In the following we will motivate our interest in investigating the consequences of imposing inappropriate restrictions on the shape of the hazard functions when modeling financial duration processes. One will recognize similarities with the conditional heteroskedasticity literature where leptokurtic unconditional return distributions triggered the development of more flexible models.

The ACD model was originally introduced for the analysis of waiting times between successive trades in intra-day markets, the so called (inter-) trade duration process. Raw trade durations are computed by $X_i = t_i - t_{i-1}$, where t_i is the time of day at which the ith trade occurred. The first data set to be analysed by means of the ACD model were trade durations for the IBM stock at the NYSE using the 1990/1991 TORQ data set (Hasbrouck1992).

A kernel density estimate and the autocorrelogram for these data are included in Figure 1. The sample selection was performed as in Engle and Russell (1998). Figure 1 depicts diurnally adjusted durations that result from dividing X_i by a time-of-day (tod) dependent function. We have adopted the standard approach by Engle and Russell (1995) and approximated the tod-function by a cubic spline regression of X_i on t_i using half hours as nodes. Duration data have a support which is bounded from below. This implies that standard density estimation methods may perform poorly due to the boundary bias that haunts fixed kernels. As a solution to this problem we have adopted the Gamma kernel approach proposed by Chen (1999) that is designed for such data.

Figure 1 depicts two features that were considered idiosyncratic for trade durations and which guided the specification of the Exponential- and the Weibull-ACD model by Engle and Russell (1998): first, the autocorrelations are significant even at higher lags; second, the density estimate takes on a right-skewed shape that resembles an exponential or Weibull distribution. This clus-

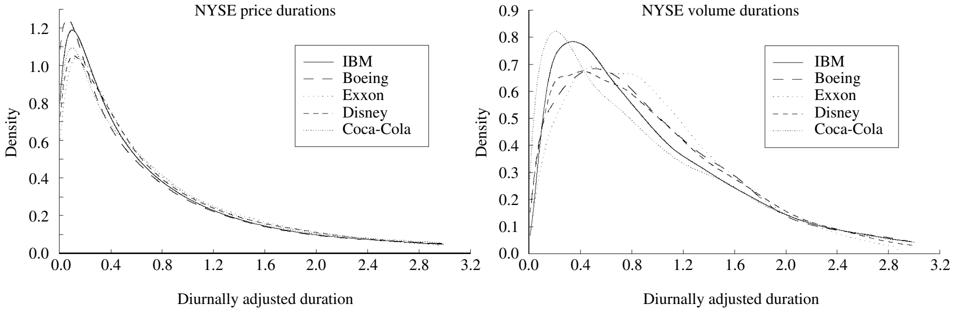

Figure 2. Density estimates using a Gamma kernel for NYSE price and volume durations (diurnally adjusted). Sample period September–November 1996, regular trading hours. See Figure 1 for bandwidth selection.

tering of small trade durations is a consequence of the data generating process and can especially be expected in electronic screen trading systems where orders are either automatically matched, or traders can initiate transactions at any time by picking off quotes displayed in the electronic order book.

Turning our attention to other financial market events it can be shown that the resulting duration distribution is of a quite different shape. Two examples are so-called price and volume durations. A price duration is defined as the time interval needed to observe a cumulative change in the mid-price not less than a threshold. The economic significance of price duration processes was outlined in the introduction. Thinning the quote process such that the selected durations are characterized by a total traded volume equal to at least v shares defines a sequence of volume durations. As pointed out by Gouriéroux *et al.* (1996), volume durations have an immediate appeal for characterizing the liquidity of a stock as they indicate the time needed to trade a given amount of shares. Whilst there is also evidence for serial correlation in volume and price durations (see Giot (1999)), the shape of the unconditional distribution is very different from that of the trade duration data. To illustrate this, Figure 2 depicts kernel density estimates for volume and price durations of five NYSE traded stocks using the Trade and Quote (TAQ) database. The event that defines a volume duration is a cumulative trading volume of at least 25 000 shares. Price durations are defined by thinning the quote process with respect to a minimum change in the mid-price of the quotes of at least \$0.125. Gamma kernel estimation and diurnal adjustment was performed as for the trade durations data.

The comparison with the trade duration density reveals striking differences. Whilst the densities are still right-skewed, they take on a clearly unimodal shape. The use of the Gamma kernel rules out the conclusion that this is a result of the boundary bias implied by fixed kernel density estimation.

In the standard ACD models proposed by Engle and Russell (1998), the hazard function conditional on past information is assumed to be either increasing, decreasing or constant with respect to duration. Let us denote by T the duration of stay in the state of interest and recall the definition of the hazard function as the instantaneous rate of leaving per unit time period at time t,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)}, \tag{1}$$

where $f(t)$ and $F(t)$ denote the duration density and distribution function, respectively. To be more precise, and in the notation of point processes employed by Engle and Russell (1998), we deal with the conditional intensity function

$$\lambda\{t \mid N(t), t_1, \ldots, t_{N(t)}\} = \lim_{\Delta t \to 0} \frac{P\{N(t+\Delta t) > N(t) \mid N(t), t_1, \ldots, t_{N(t)}\}}{\Delta t}. \tag{2}$$

In the following the expressions hazard and intensity function are used interchangeably.

Given the shape of the duration distributions in Figure 2, we raise the question whether a conditional hazard function that is first increasing and then decreasing with respect to duration might not be a more appropriate choice. Of course, the shape of the unconditional density does not necessarily imply that the restrictions concerning the conditional hazard functions in the standard ACD specifications become a practically relevant problem, but we take this as an obvious clue that it might be useful to investigate whether and when these restrictions can jeopardize the successful application of these models.

3. ECONOMETRIC MODELS

3.1. Basic models

Information events occur in clusters. This implies that the waiting times between events during the intra-day trading and quoting process exhibit a significant autocorrelation. Some of these events, for example the opening and closing of major exchanges or lunch breaks, occur with certainty. Hence, a part of the duration persistence is due to an intra-day seasonality or, more precisely, diurnality. Engle and Russell (1998) propose to decompose financial duration processes into a deterministic effect of time (also referred to as diurnal factor or time-of-day-function), $\Phi(t_i)$, and a stochastic component x_i. Defining X_i as the duration between two events that occur at t_i and t_{i-1} we have:

$$X_i = x_i \cdot \Phi(t_{i-1}). \tag{3}$$

Engle and Russell (1995) propose approximating $\Phi(t_i)$ by a regression of X_i on a spline function using polynomials of time as explanatory variables and each full hour as a node.[1] Dividing each duration in the sample by the appropriate spline function value, a sequence of diurnally adjusted durations is obtained, $x_i = \frac{X_i}{\Phi(t_{i-1})}$.

In the ACD model the conditional expected duration, $\psi_i = E(x_i \mid \mathcal{F}_i)$, where $\mathcal{F}_i$ denotes the conditioning information set generated by the durations preceding x_i, is dependent on past (expected) durations and a vector of pre-determined indicators, z_i, suggested by microstructure theory and the data generating process:

$$\psi_i = \omega + \sum_{j=1}^{q} \alpha_j x_{i-j} + \sum_{j=1}^{p} \beta_j \psi_{i-j} + \zeta z_i. \tag{4}$$

The ACD model is further characterized by the assumption that the standardized durations

$$\varepsilon_i = \frac{x_i}{f(\psi_i)}; \qquad f(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+ \tag{5}$$

[1] In this paper, we will apply the two-step method proposed by Engle and Russell (1995) in which the spline function is estimated separately from the other model parameters. Simultaneous ML estimation would also be possible. Engle and Russell (1998) report that both procedures give similar results if sufficient data are available.

are independent and identically distributed, and that their density satisfies:

$$g\left\{\frac{x_i}{f(\psi_i)}\middle|\mathcal{F}_i;\theta_g\right\} = g\left\{\frac{x_i}{f(\psi_i)};\theta_g\right\}. \tag{6}$$

This implies that the time dependence of the duration process is summarized by the conditional expected duration sequence.

In order to prevent ψ_i becoming negative, Bauwens and Giot (1997) introduced the Logarithmic ACD model in which the autoregression bears on the logarithm of the conditional expected duration. Two specifications are considered:

$$\psi_i = \exp\left\{\omega + \sum_{j=1}^{q}\alpha_j \ln(x_{i-j}) + \sum_{j=1}^{p}\beta_j \ln(\psi_{i-j}) + \zeta z_i\right\} \tag{7}$$

$$\psi_i = \exp\left\{\omega + \sum_{j=1}^{q}\alpha_j \varepsilon_{i-j} + \sum_{j=1}^{p}\beta_j \ln(\psi_{i-j}) + \zeta z_i\right\}. \tag{8}$$

Owing to its close correspondence with the EGARCH model introduced by Nelson (1991), Lunde (1999) refers to equation (8) as the Nelson form of the ACD model.

Assuming $f(\psi_i) = \psi_i$ and choosing the density in equation (6) to be the Exponential(λ) under the restriction that λ is equal to one,

$$g\left(\frac{x_i}{\psi_i}\middle|\mathcal{F}_i;\theta_g\right) = \exp\left(-\frac{x_i}{\psi_i}\right) \tag{9}$$

we obtain the Exponential-ACD (EACD). The conditional density of x_i is easily derived:

$$g(x_i \mid \mathcal{F}_i;\theta_g) = \frac{1}{\psi_i}\exp\left(-\frac{x_i}{\psi_i}\right). \tag{10}$$

This is an exponential density with distribution parameter λ equal to ψ_i^{-1}. Hence, the corresponding conditional hazard function is

$$h(x_i \mid \mathcal{F}_i;\theta_g) = \psi_i^{-1}. \tag{11}$$

Since the assumption of a constant conditional hazard function seems highly restrictive, Engle and Russell (1998) propose a more flexible alternative where

$$f(\psi_i) = \phi_i = \psi_i \cdot \left\{\Gamma\left(1+\frac{1}{\gamma}\right)\right\}^{-1} \tag{12}$$

and $g(\cdot\,;\cdot)$ is chosen to be the Weibull (λ, γ) under the restriction that $\lambda = 1$,

$$g\left(\frac{x_i}{\phi_i}\middle|\mathcal{F}_i;\theta_g\right) = \frac{\gamma\phi_i}{x_i}\left(\frac{x_i}{\phi_i}\right)^{\gamma}\exp\left\{-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right\}. \tag{13}$$

It is then straightforward to derive the conditional density of x_i,

$$g(x_i \mid \mathcal{F}_i;\theta_g) = \frac{\gamma}{x_i}\left(\frac{x_i}{\phi_i}\right)^{\gamma}\exp\left\{-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right\}, \tag{14}$$

which is a Weibull density with distribution parameter λ equal to ϕ_i. This defines the Weibull-ACD (WACD) model that reduces to the Exponential-ACD if γ is equal to one. The conditional hazard function implied by the Weibull-ACD is

$$h(x_i \mid \mathcal{F}_i; \theta_g) = \phi_i{}^{-\gamma} x_i^{\gamma-1} \gamma, \tag{15}$$

which decreases ($0 < \gamma < 1$) or increases ($\gamma > 1$) with respect to duration.

3.2. The Burr-ACD model

Due to its restrictive assumptions regarding the conditional hazard functions, the Exponential-ACD will easily be rejected in empirical applications. However, we have hypothesized that the Weibull-ACD specification still constrains the shape of the conditional hazard functions in a way that might be inappropriate when modeling financial duration processes. In the following we will propose a more flexible specification based on the Burr distribution that dates back to Burr (1942). Lancaster (1990) shows that the Burr distribution can be derived as a Gamma mixture of Weibull distributions. Exponential, Weibull and Log-Logistic are limiting cases. Unlike Weibull and Exponential, the Burr distribution is less frequently used in duration analyses. An exception is the paper by Ophem and Jonker (1996) who have considered a Burr-based model for an analysis of the duration of education spells. Since we do not assume that the reader is familiar with this distribution, the Appendix contains density, survivor and hazard functions, moments and mode, as well as useful proofs to show the limiting cases of the Burr distribution.

To derive the alternative ACD specification, we define

$$f(\psi_i) = \xi_i = \psi_i \cdot \frac{(\sigma^2)^{\left(1+\frac{1}{\kappa}\right)} \cdot \Gamma\left(\frac{1}{\sigma^2} + 1\right)}{\Gamma\left(1 + \frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)}, \tag{16}$$

where $0 < \sigma^2 < \kappa$, and choose the density in (6) to be the Burr (μ, κ, σ^2) density under the restriction that $\mu = 1$,

$$g\left(\frac{x_i}{\xi_i}\middle|\mathcal{F}_i; \theta_g\right) = \frac{\kappa \cdot \xi_i^{1-\kappa} \cdot x_i^{\kappa-1}}{(1 + \sigma^2 \cdot \xi_i^{-\kappa} \cdot x_i^{\kappa})^{\left(\frac{1}{\sigma^2}\right)+1}}. \tag{17}$$

The conditional density of x_i is then

$$g(x_i \mid \mathcal{F}_i; \theta_g) = \frac{\kappa \cdot \xi_i^{-\kappa} \cdot x_i^{\kappa-1}}{(1 + \sigma^2 \cdot \xi_i^{-\kappa} \cdot x_i^{\kappa})^{\left(\frac{1}{\sigma^2}\right)+1}}. \tag{18}$$

This is a Burr density with the μ-parameter equal to $\xi_i^{-\kappa}$. It is natural to refer to this model as the Burr-ACD (BACD). The implied conditional hazard function is

$$h(x_i \mid \mathcal{F}_i; \theta_g) = \frac{\xi_i^{-\kappa} \cdot \kappa \cdot x_i^{\kappa-1}}{1 + \sigma^2 \cdot \xi_i^{-\kappa} \cdot x_i^{\kappa}}, \tag{19}$$

which is non-monotonic with respect to duration for $\kappa \geq 1$ and $\sigma^2 > 0$. For $\sigma^2 \to 0$ the Burr-ACD reduces to the Weibull-ACD. The Exponential-ACD is contained as a special case if in addition $\kappa = 1$. Furthermore, the properties of the Burr distribution (see the Appendix) imply that for $\sigma^2 = 1$, the Burr-ACD reduces to an ACD specification that can be referred to as the Log-Logistic-ACD.

Figure 3 illustrates feasible shapes of the Burr-ACD conditional hazard functions evaluated at $\psi_i = 1$. The graphs where $\sigma^2 = 0.0001$ represent, approximately, the Weibull-ACD. The Burr-ACD log-likelihood function conditional on pre-sample values $x_{-1}, x_{-2}, \ldots$, is

$$\mathcal{L} = \sum_{i=1}^{N} \left\{ \ln \kappa - \kappa \cdot \ln \xi_i + (\kappa - 1) \cdot \ln x_i - \left(\frac{1}{\sigma^2} + 1 \right) \cdot \ln(1 + \sigma^2 \cdot \xi_i^{-\kappa} \cdot x_i^{\kappa}) \right\}. \tag{20}$$

The gradients of the Burr-ACD log-likelihood can be found in the Appendix. Optimization has to be carried out subject to the inequality constraints

$$0 < \sigma^2 < \kappa. \tag{21}$$

If specification (4) instead of (7) or (8) is chosen then further restrictions on $\omega, \alpha_1, \ldots, \alpha_q$ and $\beta_1, \ldots, \beta_q$ are required to ensure non-negativity of the conditional duration sequence. These restrictions correspond to those required for the GARCH model outlined by Nelson and Cao (1992).

4. SPECIFICATION TESTS

Despite the recent boom of empirical analyses of financial duration processes, the literature has so far devoted little attention to testing the specification of the econometric model. It is common to perform simple diagnostic tests to check whether the standardized durations are independent and identically distributed (*i.i.d.*). Whilst most papers use the Ljung-Box statistic to test for serial correlation, only a few test whether the distribution of the durations is correctly specified. In the following we review two recently proposed specification test ideas.

4.1. Testing financial duration models via density forecasts

Bauwens *et al.* (2000) propose employing the methods for evaluating density forecasts advanced by Shephard (1994), Diebold *et al.* (1997) and Kim *et al.* (1998) to test the specification of financial duration models. The motivation behind these procedures is rather intuitive and easily understood. Let us denote by $\{p_i(x_i \mid \mathcal{F}_i)\}_{i=1}^m$ a sequence of one-step-ahead density forecasts and by $\{f_i(x_i \mid \mathcal{F}_i)\}_{i=1}^m$ the sequence of densities defining the data generating process governing the duration series x_i. The one-step-ahead conditional density forecasts issued by the ACD models discussed in the previous section are given by (10), (14) and (18).

Diebold *et al.* show that the correct density is weakly superior to all other forecasts, i.e. will be preferred, in terms of expected loss, by all forecast users regardless of their loss functions. This suggests that forecasts should be evaluated by assessing whether the forecasting densities are correct, i.e. whether

$$\{p_i(x_i \mid \mathcal{F}_i)\}_{i=1}^m = \{f_i(x_i \mid \mathcal{F}_i)\}_{i=1}^m. \tag{22}$$

The distributional properties of the probability integral transform

$$z_i = \int_{-\infty}^{x_i} p_i(u)\, \mathrm{d}u \tag{23}$$

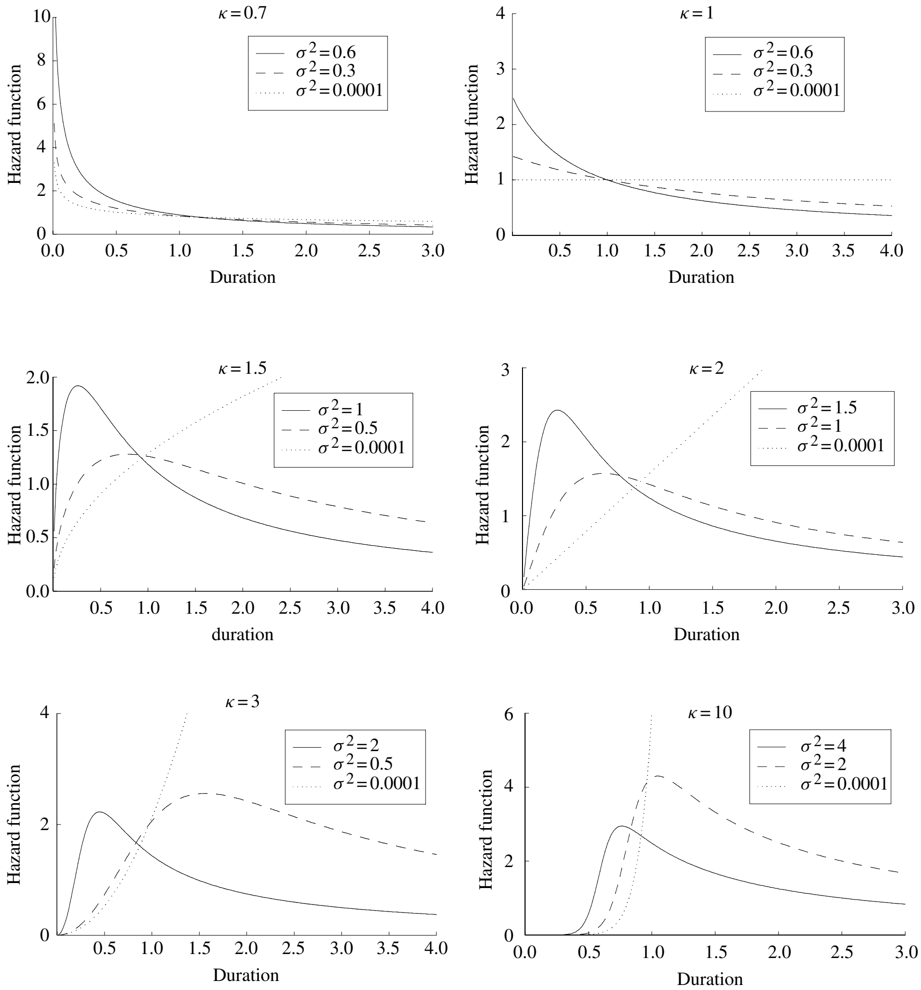

Figure 3. Burr-ACD model: feasible conditional hazard functions evaluated at $\psi_i = 1$.

derived by Rosenblatt (1952) provide the solution to the problem that $f_i(x_i \mid \mathcal{F}_i)$ is never observed. Under the null hypothesis, the distribution of the sequence of probability transforms $\{z_i\}_{i=1}^m$ of $\{x_i\}_{i=1}^m$ with respect to $\{p_i(x_i \mid \mathcal{F}_i)\}_{i=1}^m$ is *i.i.d.* $U(0, 1)$. This implies that the empirical sequence of probability integral transforms produced by the conditional duration forecasts can be used for specification testing. Diebold *et al.* (1997) recommend graphical tools that complement statistical tests for *i.i.d.* uniformity. For example, by plotting a histogram based on the empirical z-sequence, departures from uniformity can easily be detected. A straightforward χ^2 goodness-of-fit test can be computed by exploiting the statistical properties of the histogram under the null hypothesis of uniformity. Inspecting the autocorrelogram for z-sequence helps to identify potential deficiencies of a model to account for the dynamics of the duration process. Kim *et al.* (1998) used the z-sequences to look at the fit of stochastic volatility models to financial return data.

4.2. Non-parametric testing of conditional duration models

One drawback of the density evaluation method discussed in the previous subsection is that the effect of parameter estimation is not taken into account. Fernandes and Grammig (2000), henceforth referred to as F&G, introduce testing procedures that gauge the closeness between parametric and non-parametric estimates of the density functions of the standardized durations, (9), (13) and (17). To be more specific, F&G test the null

$$H_0: \qquad \exists\, \theta_g \in \Theta \qquad \text{such that} \qquad g(\cdot, \theta_g) = g(\cdot) \tag{24}$$

where $g(\cdot)$ is the true density of the standardized durations and $g(\cdot, \theta_g)$ the density implied by the parametric model. The alternative hypothesis is that there is no such $\theta_g \in \Theta$. The true density $g(\cdot)$ is of course unknown. Accordingly, F&G estimate the density function using non-parametric kernel methods, which produce consistent estimates irrespective of the parametric specification. The parametric density estimator is in turn consistent only under the null. The obvious test is therefore to gauge the closeness between these two density estimates. For that purpose, F&G consider the distance

$$\Psi_g = \int_0^\infty \mathbb{1}(x \in \mathcal{S})\{g(x, \theta) - g(x)\}^2 g(x)\, \mathrm{d}x \tag{25}$$

to build a testing procedure, which is referred to as the D-test. The compact subset $\mathcal{S}$ is introduced to avoid regions in which density estimation is unstable. The sample analog of (25) reads

$$\Psi_{\hat{g}} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(x_i \in \mathcal{S})\{g(x_i, \hat{\theta}) - \hat{g}(x_i)\}^2, \tag{26}$$

where $\hat{\theta}$ and $\hat{g}(\cdot)$ denote consistent estimates of the true parameter θ_g and density $g(\cdot)$, respectively. The null hypothesis is rejected if the D-test statistic $\Psi_{\hat{g}}$ is large enough. Under the null and a set of regularity assumptions the statistic

$$\hat{\tau}_n^D = \frac{T h_T^{1/2} \Psi_{\hat{f}} - h_T^{-1/2} \hat{\delta}_D}{\hat{\sigma}_D} \xrightarrow{d} N(0, 1). \tag{27}$$

h_T denotes the bandwidth used for the density estimation and $\hat{\delta}_D$ and $\hat{\sigma}_D^2$ are consistent estimates of $\delta_D = e_K\, E\{\mathbb{1}(x \in \mathcal{S}) f_x\}$ and $\sigma_D^2 = v_K\, E\{\mathbb{1}(x \in \mathcal{S}) f_x^3\}$, respectively, where $e_K \equiv \int_u K^2(u)\, \mathrm{d}u$ and $v_K \equiv \int_v \left\{ \int_u K(u)K(u+v)\, \mathrm{d}u \right\}^2 \mathrm{d}v$. F&G's tests are nuisance parameter free in that there is no asymptotic cost in replacing the standardized durations with their consistent estimates. All results are derived under mixing conditions, hence there is no need to perform a previous test for serial independence of the standardized durations.

The standardized durations have a support which is bounded from below. Hence, $\hat{\tau}_n^D$ may perform poorly due to the boundary bias that haunts non-parametric estimation using fixed kernels. One solution is to work with log-durations whose support is unbounded. The computation of the D-test statistic (27) based on a Gaussian kernel, log-standardized durations and the parametric densities (9), (13), (17) is straightforward using the result that for $Y = \log X$ we have $f_Y(y) = f_X\{\exp(y)\} \exp(y)$.

Table 1. Data generating processes. R denotes the number of replications and N the sample size.

	ω	α	β	γ	κ	σ^2	R	N
DGP I	0.20	0.10	0.70	—	—	—	1000	15 000
DGP II	0.01	0.02	0.97	—	—	—	1000	15 000
DGP III	0.20	0.10	0.70	0.60	—	—	1000	15 000
DGP IV	0.01	0.02	0.97	0.60	—	—	1000	15 000
DGP V	0.20	0.10	0.70	—	2.00	1.50	1000	15 000
DGP VI	0.01	0.02	0.97	—	2.00	1.50	1000	15 000

5. MONTE CARLO STUDY

In this section we address two questions. First, we assess how severe the consequences of erroneously imposing monotonic conditional hazard functions in ACD models are. We are especially interested in investigating whether the ML estimators for the parameters $\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p$, which are required to predict conditional expected durations, are affected by a misspecification of the conditional hazard function. This is a crucial issue especially for the class of ACD-GARCH models where the ACD model is employed to predict expected durations which enter the volatility equation in the form of explanatory variables. Second, we investigate whether the inequality constraints (21) cause problems for ML estimation of the Burr-ACD.

We focus on six different data generating processes (DGP) based on the autoregression (4). DGP I and II are Exponential-ACD(1,1) processes. The parameters ω, α and β are chosen so that they crudely mimic those found in empirical applications, and so that the unconditional expected duration equals one. Note that for an ACD(1,1) we have $E(x_i) = \frac{\omega}{1-\alpha-\beta}$. DGP III and IV are Weibull-ACD processes where $\gamma = 0.6$. DGP V and VI are duration processes that imply non-monotonic hazard functions with durations created by Burr-ACD processes, where $\kappa = 2$ and $\sigma^2 = 1.5$. The simulation study is based on $R = 1000$ replications and a sample size of $N = 15\,000$ for each DGP. Table 1 summarizes the design.

After each replication, ML estimations are carried out using the GAUSS programs for the computation of the log-likelihood functions and gradients for Exponential-, Weibull-, and Burr-ACD written by the authors. We apply the sequential quadratic programming algorithm (SQP) for the optimization of functions with general inequality and equality constraints proposed by Han (1977). The SQP algorithm is a good choice since all ACD specifications based on the autoregression (4) impose non-negativity constraints on expected durations. The Burr-ACD additionally imposes the inequality constraints (21). Detailed results are reported in Table 5 that is deferred to the Appendix. In addition to the quantiles, mean and standard deviation of the estimators, the root mean squared error (RMSE) and the mean of the absolute error (MAE) are computed as measures for the accuracy of the estimates. Figure 4 additionally depicts kernel density plots.

Table 5 shows that the Burr-ACD easily reduces to either Exponential- or Weibull-ACD, as is required for DGPs I–IV. The Burr-ACD ML estimators are as precise as the ML estimators that correspond to the true DGP. For the majority of replications $\hat{\sigma}_r^2$ converges to the lower bound (1×10^{-5}), and is not far above for the rest. The distribution of $\hat{\kappa}_r$ is narrowly centered around one for DGP I and II, and is very similar to the distribution of $\hat{\gamma}_r$ for DGP III and IV. Hence, the

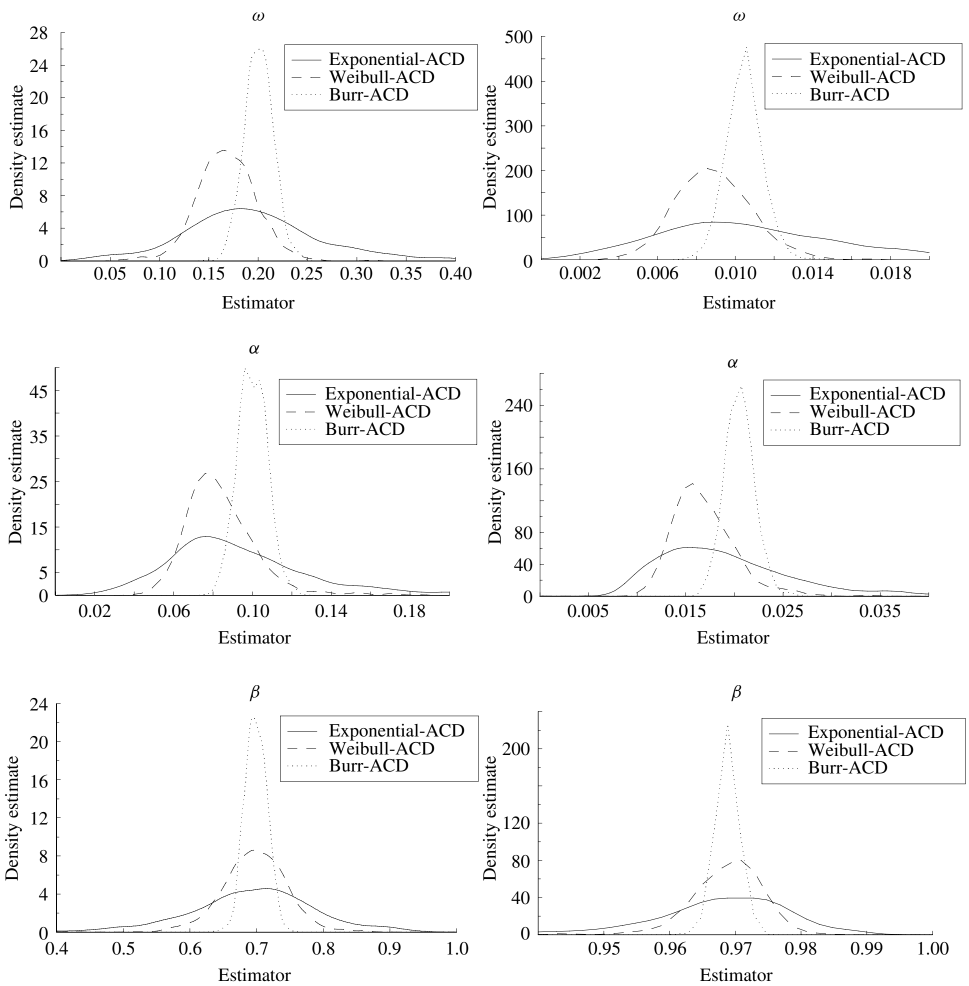

Figure 4. Kernel plots of ACD ML estimators DGP V (left) and VI (right). Gaussian kernel with bandwidth $h = (0.9s \cdot N)^{-0.2}$, where s is the standard deviation and N the number of observations in the sample.

inequality restrictions (21) do not affect the performance of the ML estimation.

The results for DGP V and VI reveal that a misspecification of the hazard function does indeed entail severe consequences. RMSE and MAE that are produced by the Exponential- and Weibull-ACD QML estimators are large. The kernel plots in Figure 4 show the low relative efficiency of the Exponential- and Weibull-ACD ML estimators for β. Even worse, the Weibull-ACD tends to underestimate the true parameters ω and α. The Exponential-ACD ML estimators are even more inefficient, but the bias is less pronounced. Despite the fact that the distributions of the Exponential-ACD ML estimators appear somewhat right-skewed, the means of the estimators $\hat{\omega}_r$ and $\hat{\alpha}_r$ are closer to the true values.

The conclusion that the misspecification of the conditional hazard functions can result in serious problems for predicting expected durations is emphasized when computing the uncon-

Table 2. Descriptive statistics of price durations. Overdispersion stands for the ratio between standard deviation and mean. $Q(10)$ denotes the Ljung-Box statistic of order 10.

Stock	Sample size	Mean	Overdispersion	$Q(10)$
Boeing (BA)	2620	1.001	1.338	322.3
Coca-Cola (KO)	1609	1.002	1.171	69.7
Disney (DIS)	2160	1.004	1.209	137.3
Exxon (XON)	2717	1.000	1.196	68.2
IBM	6728	1.015	1.427	1932.6

ditional expected duration estimate that is implied by the mean of the estimates $\hat{\omega}_r$, $\hat{\alpha}_r$ and $\hat{\beta}_r$. Recall that the true parameters were chosen such that the unconditional expected duration equals one. For DGP V (DGP VI) the Weibull-ACD implies that $\hat{E}(x_i)$ is equal to 0.77 (0.62). The Exponential-ACD performs better with $\hat{E}(x_i) = 0.97$ (DGP V) and $\hat{E}(x_i) = 1.05$ (DGP VI), but the price is the poor efficiency of the ML estimator.

6. EMPIRICAL APPLICATION

6.1. Data

In this section we use real world data to assess the significance of allowing for non-monotonic hazard functions when modeling financial duration processes. We restrict our attention to the modeling of price durations of NYSE traded stocks. Data were provided by Luc Bauwens and Pierre Giot, who have constructed a powerful database from the NYSE Trade and Quote (TAQ) raw data. We use data ranging from September to November 1996 and analyse price durations of five actively traded stocks: Boeing, Coca-Cola, Disney, Exxon, and IBM. Trading at the NYSE is organized as a combined market maker/order book system. A designated specialist composes the market for each stock by managing the trading and quoting processes and providing liquidity. Apart from an opening auction, trading is continuous from 9:30 to 16:00. We define a price duration as the time interval needed to observe a cumulative change in the mid-price of at least \$0.125. For all stocks, durations between events recorded outside the regular opening hours of the NYSE as well as inter-day durations are removed. As documented by Giot (1999), price durations feature a strong diurnality related to pre-determined market characteristics such as trade opening and closing times and lunch times. Diurnally adjusted durations are computed as outlined in Section 3.1. Cubic splines using half hour nodes are employed to smooth the time-of-day function $\Phi(t_i)$. Separate splines are estimated for each day of the week. Table 2 reports descriptive statistics of the resulting price duration sequences. The two common features across stocks are a highly significant serial correlation and some degree of overdispersion.

6.2. Estimation and test results

For each stock we used the first two-thirds of the observations for estimation and in-sample testing. The remainder was reserved for out-of-sample testing. We restricted our attention to a parsimonious ACD(1,1) specification using the autoregression (4) and did not include pre-determined

Table 3. Parameter estimates and robust standard errors.

		ω	α	β	κ	σ^2	$\mathcal{L}$
	EACD	0.031	0.114	0.861			−1784.7
		(0.023)	(0.041)	(0.059)			
BA	WACD	0.034	0.121	0.851	0.895		−1764.4
		(0.025)	(0.042)	(0.061)	(0.016)		
	BACD	0.057	0.169	0.789	1.093	0.339	−1740.1
		(0.033)	(0.046)	(0.067)	(0.036)	(0.061)	
	EACD	0.159	0.109	0.727			−1016.5
		(0.042)	(0.026)	(0.051)			
KO	WACD	0.159	0.109	0.727	0.959		−1014.8
		(0.042)	(0.026)	(0.051)	(0.019)		
	BACD	0.161	0.124	0.715	1.124	0.286	−1007.1
		(0.042)	(0.030)	(0.051)	(0.050)	(0.079)	
	EACD	0.074	0.046	0.889			−1613.0
		(0.030)	(0.015)	(0.033)			
DIS	WACD	0.074	0.046	0.888	0.969		−1611.8
		(0.031)	(0.015)	(0.034)	(0.018)		
	BACD	0.099	0.048	0.867	1.219	0.396	−1588.0
		(0.044)	(0.018)	(0.049)	(0.045)	(0.067)	
	EACD	0.065	0.046	0.890			−1803.2
		(0.037)	(0.016)	(0.048)			
XON	WACD	0.066	0.045	0.889	0.962		−1800.8
		(0.038)	(0.016)	(0.049)	(0.016)		
	BACD	0.102	0.039	0.863	1.250	0.464	−1766.2
		(0.055)	(0.015)	(0.061)	(0.044)	(0.068)	
	EACD	0.010	0.090	0.905			−5044.3
		(0.005)	(0.019)	(0.021)			
IBM	WACD	0.010	0.090	0.904	0.985		−5043.4
		(0.005)	(0.019)	(0.021)	(0.011)		
	BACD	0.017	0.112	0.880	1.263	0.420	−4952.0
		(0.009)	(0.029)	(0.033)	(0.025)	(0.038)	

explanatory variables. We have also estimated Logarithmic-ACD versions of the Exponential-, Weibull- and Burr-ACD based on (7) and (8), but the results do not differ qualitatively. Maximum Likelihood estimates and robust standard errors are reported in Table 3.

For most stocks, the Exponential- and Weibull-ACD estimates of α and β are quite similar, whereas the Burr-ACD estimates differ considerably. All $\hat{\gamma}$ are less than and close to one which implies monotonically decreasing Weibull-ACD hazard functions. Furthermore, the log-likelihood values of Exponential- and Weibull-ACD are quite close, whereas the Burr-ACD produces considerably higher log-likelihood values. Using a standard likelihood ratio statistic, Exponential- and Weibull-ACD are clearly to be rejected in favor of the Burr-ACD model. All $\hat{\kappa}$ are significantly greater than one, and the $\hat{\sigma}^2$ are clearly above the zero lower bound. This implies that the Burr-ACD conditional hazard functions are non-monotonic for all stocks. Figure 5 illustrates this result and depicts the non-monotonic standardized duration densities and the conditional hazard functions evaluated at $\psi_i = 1$ implied by Burr-ACD ML estimates.

Table 4 contains the numerical output of the diagnostic procedures outlined in Section 4. The D-test statistics are computed based on a Gaussian kernel and log-durations. With the sole exception of Coca-Cola, Exponential- and Weibull-ACD are clearly rejected. In contrast, the Burr-ACD p-values are quite large. IBM is an exception, since all models are rejected, but the Burr-ACD produces the best results.

The results of the χ^2 goodness-of-fit test designed for evaluating the models' density forecasts are in line with the D-test results. Exponential- and Weibull-ACD are clearly rejected for all stocks, including Coca-Cola. Again, the Burr-ACD produces large p-values for Boeing, Coca-Cola, and Disney. The p-values for Exxon and IBM are smaller, but the superior performance compared to Exponential- and Weibull-ACD is obvious.

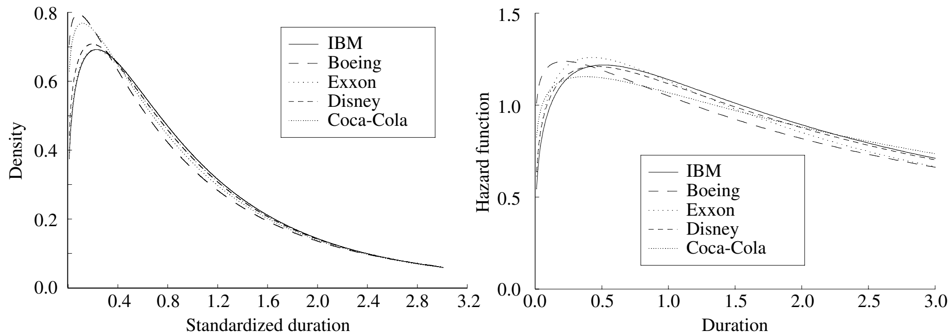

Figure 5. Standardized duration density and conditional hazard functions ($\psi = 1$) implied by Burr-ACD ML estimates.

Table 4. Specification test results.

		D-Test		χ^2−GOF		AC(z)	
Stock		In	Out	In	Out	In	Out
	EACD	0.00	0.00	0.00	0.00	8	4
BA	WACD	0.00	0.00	0.00	0.00	9	4
	BACD	13.77	0.94	20.10	0.01	7	4
	EACD	2.91	82.07	0.00	1.77	3	0
KO	WACD	31.62	87.68	0.02	11.40	2	0
	BACD	66.57	96.89	19.72	6.15	2	1
	EACD	0.00	0.00	0.00	0.00	5	21
DIS	WACD	0.00	0.00	0.00	0.00	1	21
	BACD	15.98	0.00	5.10	0.00	4	25
	EACD	0.00	0.67	0.00	0.01	5	2
XON	WACD	0.00	2.80	0.00	0.00	5	2
	BACD	13.71	26.06	2.56	5.52	3	2
	EACD	0.00	0.00	0.00	0.00	8	4
IBM	WACD	0.00	0.00	0.00	0.00	7	4
	BACD	0.28	0.00	0.03	0.00	6	6

An advantage of our specification tests is that they allow for visual diagnostic checks that are helpful for interpreting the numerical results. We focus on the EXXON and IBM results, representing the best and worst case. The graphs for the other stocks look similar. The upper panels in Figures 5 and 6 depict the non-parametric density estimates for the log-standardized durations together with their parametric counterparts as implied by Weibull- and Burr-ACD. The two panels in the middle of the figures display the histogram and the two lower panels the autocorrelograms of the sequence of probability integral transforms. The charts for the

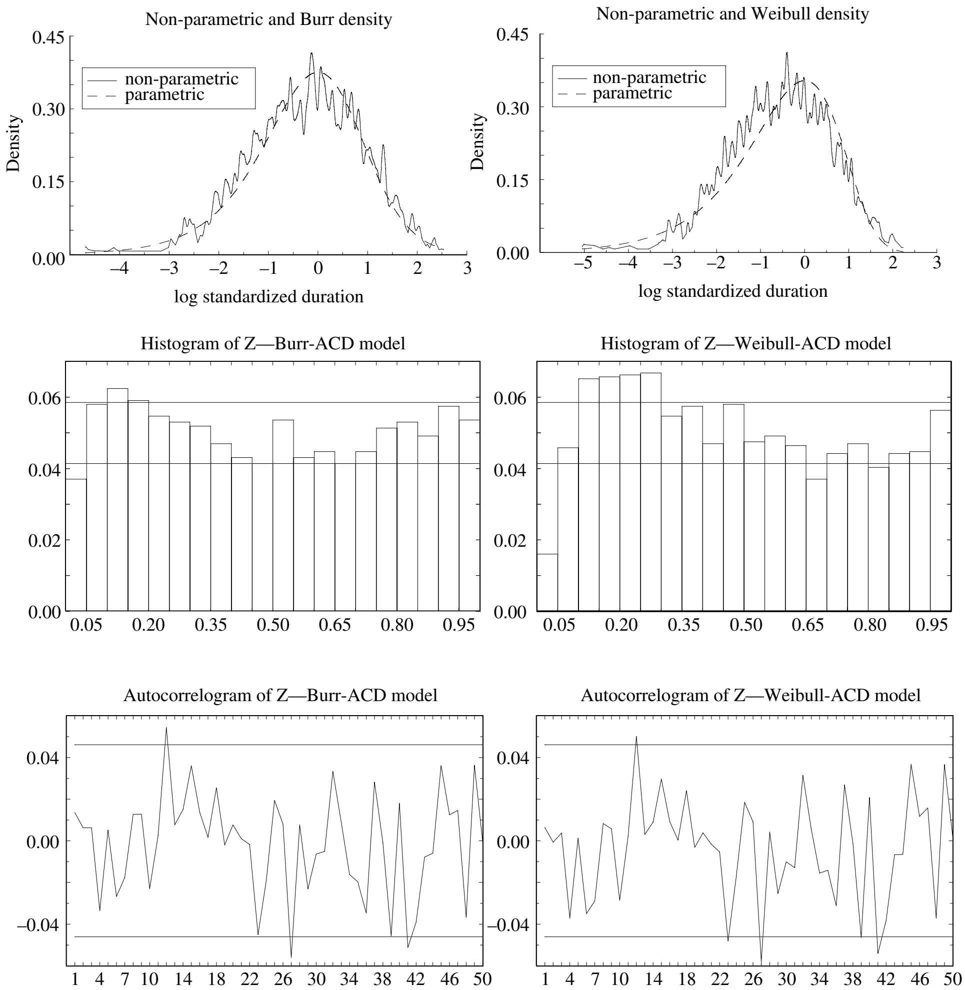

Figure 6. Visual diagnostic checks: Exxon in-sample results.

Exponential-ACD are worse than those for the Weibull-ACD. Hence, we restrict our attention to a comparison of the latter and the Burr-ACD.

The density plots in the upper panels provide graphical intuition for the D-test results. The non-parametric densities appear to fluctuate tightly around the Burr-ACD parametric densities, whereas the parametric densities implied by the Weibull-ACD remain for large intervals consistently below their non-parametric counterparts. In accordance with this result the z-histograms show clearly the superiority of the Burr-ACD, since the bars remain well within the limits of the 90% confidence intervals. Note that this also holds true for the IBM stock where, although the formal tests rejected the Burr-ACD at 1% significance level, the improved performance compared with the Weibull-ACD is unquestionable with respect to both D-test and density forecast. Deviations from uniformity are striking for the Weibull-ACD that evidently finds it difficult to account for the durations at the lower bound of the distribution.

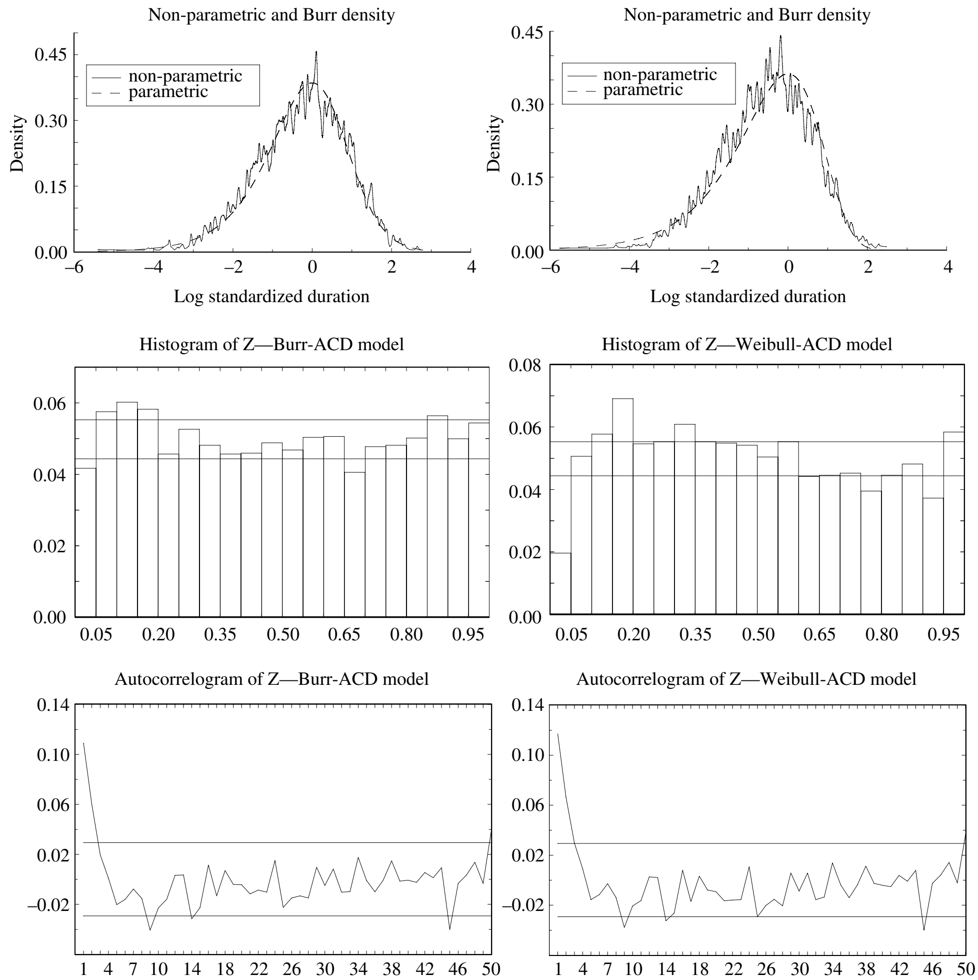

Figure 7. Visual diagnostic checks: IBM in-sample results.

All ACD specifications capture the duration dynamics in more or less the same way (with slight advantages for the Burr-ACD). The last two columns in Table 4 contains the number of autocorrelations (out of 50) for z that are significant at the 5% level. The lower panels of Figures 5 and 6 identify the lags at which significant autocorrelations occur. With the sole exception of IBM, the results can be considered satisfactory for all stocks.

Table 4 also reports the out-of-sample evaluation and test results. This represents a much tougher exercise since the parameters were estimated on the first two-thirds of the sample, but the test statistics were computed on the basis of the last one-third of the sample. Nevertheless, the superior performance of the Burr-ACD is indisputable.

7. SUMMARY AND CONCLUSION

In this paper we have argued for the necessity of allowing for non-monotonic hazard functions in the class of Autoregressive Conditional Duration models. Starting from a descriptive analysis of empirical volume and price duration processes, we have hypothesized that the assumption of monotonic hazard functions that is maintained in the standard ACD specifications might be too restrictive. We have introduced a more flexible alternative that relaxes these assumptions, and contains the basic ACD specifications proposed by Engle and Russell (1998) and others that have not yet been considered in the literature as special cases.

In a simulation study we analysed the consequences of a misspecification of the conditional hazard functions. We found that the QML estimators of the standard ACD models tend to be biased and inefficient when the true DGP requires non-monotonic hazard functions. The crucial point is that bias and inefficiency also affect the estimators of the parameters that are needed to predict expected durations. This has severe consequences for the class of ACD-GARCH models recently introduced by Engle (2000) and Ghysels and Jasiak (1998b) in which ACD models are employed to predict conditional expected durations that enter the conditional heteroskedasticity equation in the form of explanatory variables.

In an empirical application we have shown that allowing for non-monotonic hazard functions is an important issue for modeling price duration processes of NYSE traded stocks. Employing both a recently proposed specification test and the density forecast evaluation techniques, we found that the standard ACD specifications are clearly rejected, whereas the model introduced in this paper delivered unmistakably superior results. We conclude that the hazard function flexibility that is offered by the new specification turns out to be a crucial success factor.

ACKNOWLEDGEMENTS

We thank the audiences at the 53rd Econometric Society European Meeting, at the 9th $(EC)^2$ conference, and seminar participants at CORE/Université Catholique de Louvain and the University of Mannheim for helpful comments. We have benefited from the insights of T. Amemiya, L. Bauwens, R. Engle, M. Fernandes, R. Hujer, P. Giot, W. Härdle, S. Kokot, J. Krahnen, C. MacKinlay, S. Mittnik, E. Theissen, and D. Veredas. The comments and suggestions of the editor and an anonymous referee helped greatly to improve the quality of this paper. We are especially grateful to Luc Bauwens and Pierre Giot for providing access to the NYSE price durations data used in the empirical application. Financial support from the BMW foundation, the Landeszentralbank Hessen and Roland Berger & Partners is gratefully acknowledged. We retain the responsibility for any remaining errors.

REFERENCES

Bauwens, E. and P. Giot (1998). Asymmetric ACD models: Introducing price information in ACD models with a two state transition model. CORE Discussion Paper 9844, Université Catholique de Louvain.

Bauwens, L. and P. Giot (1997). *The Logarithmic ACD Model: An Application to Market Microstructure and NASDAQ*. Université Catholique de Louvain.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas (2000). *A Comparison of Financial Duration Models via Density Forecasts*. Université Catholique de Louvain and University of Frankfurt.

Bauwens, L. and D. Veredas (1999). *The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations*. Université Catholique de Louvain.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 307–7.

Burr, I. W. (1942). Cumulative fequency functions. *Annals of Mathematical Statistics 13*, 215–32.

Chen, S. X. (1999). *Probability Density Function Estimation using Gamma Kernels*. School of Statistical Science, La Trobe University.

Cox, D. R. (1955). Some statistical models connected with series of events (with discussion). *Journal of the Royal Statistical Society, Series B 34*, 129–64.

Cox, D. R. and P. A. W. Lewis (1966). *The Statistical Analysis of Series of Events*. London, New York: Wiley.

Diamond, D. W. and R. E. Verrecchia (1987). Constraints on short-selling and asset price adjustments to private information. *Journal of Financial Economics 18*, 227–311.

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

Easley, D., N. Kiefer, M. O'Hara, and J. P. Paperman (1996). Liquidity, information and infrequently traded stocks. *Journal of Finance 51*, 1405–36.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. (2000). The econometrics of ultra high frequency data. *Econometrica 68*, 1–22.

Engle, R. F. and A. Lunde (1998). Trades and quotes: A bivariate point process. Economics Discussion Paper 9807, University of Califonia at San Diego.

Engle, R. F. and J. R. Russell (1995). Autoregressive conditional duration: A new model for irregularly spaced time series data. Discussion Paper 9427, University of California.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: A new model for irregulary spaced transaction data. *Econometrica 66*, 1127–62.

Fernandes, M. and J. Grammig (2000). *Non-parametric Specification Tests for Conditional Duration Models*. European University Institute and University of Frankfurt.

Gaver, D. P. and P. A. W. Lewis (1980). First order autoregressive gamma sequences and point processes. *Advances in Applied Probability 12*, 727–45.

Gerhard, F. and N. Hautsch (1999). Volatility estimation on the basis of price intensities. Technical Report 19, Center of Finance and Econometrics (COFE), University of Konstanz.

Ghysels, E., C. Gouriéroux, and J. Jasiak (1998). Stochastic volatility duration models. Working paper 9746 CREST, Paris.

Ghysels, E. and J. Jasiak (1998a). GARCH for irregularly spaced financial data: The ACD-GARCH model. *Studies in Nonlinear Economics and Econometrics 2*, 133–49.

Ghysels, E. and J. Jasiak (1998b). Long-term dependence in trading. Discussion Paper, Penn State University and York University.

Giot, P. (1999). Time transformations, intraday data, and volatility models. CORE Discussion Paper 9944, Université catholique de Louvain, Louvain-La-Neuve.

Glosten, L. and P. Milgrom (1985). Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 1*, 71–100.

Gouriéroux, C., J. Jasiak, and G. Le Fol (1996). Intraday market activity. CREST Working Paper 9633.

Hamilton, J. D. and O. Jorda (1999). A model for the federal funds rate target. Working Paper 99-07, University of California at Davis.

Han, S. P. (1977). A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications 22*, 297–309.

Hasbrouck, J. (1988). Trades, quotes, inventories and information. *Journal of Finance 46*, 179–252.

Hasbrouck, J. (1992). Using the TORQ database. Working Paper.

Hawkes, A. G. (1971a). Spectra of some mutually exciting point processes with associated variables. *Journal of the Royal Statistical Society 33*, 438–43.

Hawkes, A. G. (1971b). Spectra of some self-exciting and mutually exciting point processes. *Biometrika 58*, 83–90.
Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In Lewis, P. A. W. (Ed.), *Stochastic Point Processes*, pp. 261–71. Wiley.
Jacobs, P. A. and P. A. W. Lewis (1977). A mixed autoregressive-moving average exponential sequence and point process (earma1,1). *Advances in Applied Probability 9*, 87–94.
Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–93.
Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
Lawrence, A. J. and P. A. W. Lewis (1980). The exponential autoregessive-moving average EARMA(p,q) process. *Journal of the Royal Statistical Society 42*, 150-61.
Lunde, A. (1999). *A Generalized Gamma Autoregressive Conditional Duration Model*. Aalborg University.
Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*, 347–70.
Nelson, D. B. and C. Q. Cao (1992). Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics 10*, 229–35.
Ogata, Y. and K. Katsura (1986). Point-processes models with linearly parameterized intensity for the application of earthquake catalogue. *Journal of Applied Probability 23*, 231–40.
O'Hara, M. (1995). *Market Microstructure Theory*. Oxford: Basil Blackwell.
Ophem, H. V. and N. Jonker (1996). The duration of higher education: A dependent competing risks model. Working Paper, University of Amsterdam.
Pringent, J.-L., E. Renault, and O. Scaillet (1999). *An Autoregressive Conditional Binomial Option Pricing Model*. Université de Cergy-Pontoise, CREST, and Université Catholique de Louvain.
Rosenblatt, R. F. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–2.
Rubin, I. (1972). Regular point processes and their detection. In ITT-18 (Ed.), *IEEE Transactions on Information Theory*, 547–57.
Russell, J. R. (1998). *Econometric Modeling of Multivariate Irregularly-spaced High-frequency Data*. Graduate School of Business, University of Chicago.
Russell, J. R. and R. F. Engle (1998). Econometric analysis of discrete-valued irregularly-spaced financial transaction data using a new autoregressive conditional multinomial model. Discussion Paper 9810, University of California at San Diego.
Shephard, N. (1994). Partial non-gaussian state space. *Biometrika 81*, 115–31.
Vere-Jones, D. and T. Ozaki (1982). Some examples of statistical inference applied to earthquake data. *Annals of the Institute for Statistical Mathematics 34*, 289–207.
Wold, H. (1948). On stationary point processes and markov chains. *Skandinavisk Aktuarietidskrist 31*, 229–40.
Zhang, M. Y., J. R. Russell, and R. S. Tsay (1999). *A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data*. Graduate School of Business, University of Chicago.

APPENDIX

A.1. Properties of the Burr distribution

Lancaster (1990) derives the Burr distribution as a Gamma mixture of Weibull distributions. Density, hazard and survivor functions of random variable T are given by

$$f(t) = \frac{\mu\kappa t^{\kappa-1}}{(1+\sigma^2\mu t^{\kappa})^{\frac{1}{\sigma^2}+1}} \tag{A.1}$$

$$S(t) = (1 + \sigma^2 \mu t^\kappa)^{-\frac{1}{\sigma^2}} \tag{A.2}$$

$$h(t) = \frac{\mu \kappa t^{\kappa - 1}}{1 + \sigma^2 \mu t^\kappa}. \tag{A.3}$$

For $\kappa \geq 1$ the mode of the distribution is equal to $\left\{\frac{\kappa-1}{\mu(\sigma^2+\kappa)}\right\}^{\frac{1}{\kappa}}$. The jth moment of T is

$$E(T^j) = \mu^{-\frac{j}{\kappa}} \frac{\Gamma\left(1 + \frac{j}{\kappa}\right)\Gamma\left(\frac{1}{\sigma^2} - \frac{j}{\kappa}\right)}{\sigma^{2\left(1+\frac{j}{\kappa}\right)}\Gamma\left(\frac{1}{\sigma^2} + 1\right)}. \tag{A.4}$$

Deriving of the Burr-ACD we have stated that the Burr distribution approximates the Weibull distribution for $\sigma^2 \to 0$ approaching zero. This can be seen by rewriting the Burr density (A.1) as

$$f(t) = \frac{\kappa t^{-1}\left(\mu^{\frac{1}{\kappa}} t\right)^\kappa}{\left\{1 + \sigma^2\left(\mu^{\frac{1}{\kappa}} t\right)^\kappa\right\}^{\frac{1}{\sigma^2+1}}}. \tag{A.5}$$

Making use of the result that

$$\begin{aligned} \lim_{h \to 0}\left(1 + \frac{h}{x}\right)^{\frac{1}{h}+1} &= \lim_{h \to 0}\left[\exp\left\{\left(\frac{1}{h} + 1\right)\ln\left(1 + \frac{h}{x}\right)\right\}\right] \\ &= \exp\left\{\lim_{h \to 0} \frac{\ln(x+h) - \ln(x)}{h}\right\} \\ &= \exp\left(\frac{1}{x}\right) \end{aligned} \tag{A.6}$$

and taking the limit of (A.5) for $\sigma^2 \to 0$ we have

$$\lim_{\sigma^2 \to 0} f(t) = \frac{\kappa}{t}(\mu^{-\kappa} t)^\kappa \cdot \exp\{-(\mu^{-\kappa} t)^\kappa\}. \tag{A.7}$$

Recall that density, hazard and survivor functions of the Weibull distribution are given by

$$f(t) = \lambda\gamma(\lambda t)^{\gamma - 1} \tag{A.8}$$

$$S(t) = \exp\{-(\lambda t)^\gamma\} \tag{A.9}$$

$$h(t) = \lambda\gamma(\lambda t)^{\gamma - 1}. \tag{A.10}$$

Thus, for $\sigma^2 \to 0$ the Burr reduces to the Weibull distribution with parameters $\lambda = \mu^{-\kappa}$ and $\kappa = \gamma$. If in addition κ is equal to one, then (A.7) reduces to the Exponential density.

Comparing (A.2) to the survivor function of the Log-Logistic given by

$$S(t) = \frac{1}{1 + (\lambda t)^\kappa} \tag{A.11}$$

it is readily seen that the Burr contains the Log-Logistic distribution as a special case for $\sigma^2 = 1$ and λ in (A.11) equal to $\mu^{-\kappa}$.

A.2. First-order conditions for the maximum of the Burr-ACD log-likelihood function

$$\frac{\partial \mathcal{L}}{\partial \kappa} = \sum_{i=1}^{N} \Bigg[\frac{1}{\kappa} - \frac{1}{\kappa} \cdot \frac{1}{\xi_i} \cdot \frac{\partial \xi_i}{\partial \kappa} + \ln x_i - \left\{ \left(\frac{1}{\sigma^2} \right)^2 + 1 \right\} \cdot \frac{\sigma^2}{1 + \sigma^2 \xi_i^{-\kappa} x_i^{\kappa}} \cdot \left(\frac{x_i}{\xi_i} \right)^{\kappa} \left\{ \ln \left(\frac{x_i}{\xi_i} \right) - \frac{\kappa}{\xi_i} \frac{\partial \xi_i}{\partial \kappa} \right\} \Bigg]$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = \sum_{i=1}^{N} \Bigg[-\frac{\kappa}{\xi_i} \cdot \frac{\partial \xi_i}{\partial \sigma^2} - \left\{ 1 - \left(\frac{1}{\sigma^2} \right)^2 \right\} \cdot \ln \left[1 + \sigma^2 \cdot \xi_i^{-\kappa} x_i^{\kappa} \right] - \frac{\left(\frac{1}{\sigma^2} + 1 \right)}{1 + \sigma^2 \xi_i^{-\kappa} x_i^{\kappa}} \left(\frac{x_i}{\xi_i} \right)^{\kappa} \left(1 - \sigma^2 \frac{\kappa}{\xi_i} \cdot \frac{\partial \xi_i}{\partial \sigma^2} \right) \Bigg]$$

$$\frac{\partial \mathcal{L}}{\partial \omega} = \sum_{i=1}^{N} \left(\left[-\frac{\kappa}{\xi_i} - \left(\frac{1}{\sigma^2} + 1 \right) \cdot \frac{1}{1 + \sigma^2 \xi_i^{-\kappa} x_i^{\kappa}} \cdot \left\{ -\sigma^2 \cdot \left(\frac{x_i}{\xi_i} \right)^{\kappa} \cdot \frac{\kappa}{\xi_i} \right\} \right] \cdot \frac{\partial \xi_i}{\partial \omega} \right)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_r} = \sum_{i=1}^{N} \left(\left[-\frac{\kappa}{\xi_i} - \left(\frac{1}{\sigma^2} + 1 \right) \cdot \frac{1}{1 + \sigma^2 \xi_i^{-\kappa} x_i^{\kappa}} \cdot \left\{ -\sigma^2 \cdot \left(\frac{x_i}{\xi_i} \right)^{\kappa} \cdot \frac{\kappa}{\xi_i} \right\} \right] \cdot \frac{\partial \xi_i}{\partial \alpha_r} \right)$$

$$\frac{\partial \mathcal{L}}{\partial \beta_r} = \sum_{i=1}^{N} \left(\left[-\frac{\kappa}{\xi_i} - \left(\frac{1}{\sigma^2} + 1 \right) \cdot \frac{1}{1 + \sigma^2 \xi_i^{-\kappa} x_i^{\kappa}} \cdot \left\{ -\sigma^2 \cdot \left(\frac{x_i}{\xi_i} \right)^{\kappa} \cdot \frac{\kappa}{\xi_i} \right\} \right] \cdot \frac{\partial \xi_i}{\partial \beta_r} \right)$$

$$\frac{\partial \mathcal{L}}{\partial \zeta_r} = \sum_{i=1}^{N} \left(\left[-\frac{\kappa}{\xi_i} - \left(\frac{1}{\sigma^2} + 1 \right) \cdot \frac{1}{1 + \sigma^2 \xi_i^{-\kappa} x_i^{\kappa}} \cdot \left\{ -\sigma^2 \cdot \left(\frac{x_i}{\xi_i} \right)^{\kappa} \cdot \frac{\kappa}{\xi_i} \right\} \right] \cdot \frac{\partial \xi_i}{\partial \zeta_r} \right)$$

where

$$\frac{\partial \xi_i}{\partial \kappa} = \psi_i \cdot \frac{-\sigma^{2\left(1+\frac{1}{\kappa}\right)} \cdot \Gamma\left(\frac{1}{\sigma^2} + 1\right) \cdot \left\{ \ln \sigma^2 + \Gamma'\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right) - \Gamma'\left(1 + \frac{1}{\kappa}\right) \right\}}{\kappa^2 \cdot \Gamma\left(1 + \frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)}$$

$$\frac{\partial \xi_i}{\partial \sigma^2} = \psi_i \cdot \frac{\sigma^{\frac{2}{\kappa}} \cdot \Gamma\left(\frac{1}{\sigma^2} + 1\right) \cdot \left[\left(1 + \frac{1}{\kappa}\right) + \frac{1}{\sigma^2} \cdot \left\{ \Gamma'\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right) - \Gamma'(\frac{1}{\sigma^2} + 1) \right\} \right]}{\Gamma\left(1 + \frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)}$$

$$\frac{\partial \xi_i}{\partial \omega} = \left\{ \frac{\sigma^{2\left(1+\frac{1}{\kappa}\right)} \cdot \Gamma\left(\frac{1}{\sigma^2} + 1\right)}{\Gamma\left(1 + \frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)} + \beta_1 \cdot \frac{\partial \xi_{i-1}}{\partial \omega} + \cdots + \beta_p \cdot \frac{\partial \xi_{i-p}}{\partial \omega} \right\}$$

$$\frac{\partial \xi_i}{\partial \alpha_r} = \left\{ x_{i-r} \cdot \frac{\sigma^{2\left(1+\frac{1}{\kappa}\right)} \cdot \Gamma\left(\frac{1}{\sigma^2} + 1\right)}{\Gamma\left(1 + \frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2} - \frac{1}{\kappa}\right)} + \beta_1 \cdot \frac{\partial \xi_{i-1}}{\partial \alpha_r} + \cdots + \beta_p \cdot \frac{\partial \xi_{i-p}}{\partial \alpha_r} \right\}$$

$$\frac{\partial \xi_i}{\partial \beta_r} = \left(\xi_{i-r} + \beta_1 \cdot \frac{\partial \xi_{i-1}}{\partial \beta_r} + \cdots + \beta_p \cdot \frac{\partial \xi_{i-p}}{\partial \beta_r} \right)$$

$$\frac{\partial \xi_i}{\partial \zeta_r} = \left\{ z_r \cdot \frac{\sigma^{2\left(1+\frac{1}{\kappa}\right)} \cdot \Gamma\left(\frac{1}{\sigma^2}+1\right)}{\Gamma\left(1+\frac{1}{\kappa}\right) \cdot \Gamma\left(\frac{1}{\sigma^2}-\frac{1}{\kappa}\right)} + \beta_1 \cdot \frac{\partial \xi_{i-1}}{\partial \delta_r} + \cdots + \beta_p \cdot \frac{\partial \xi_{i-p}}{\partial \delta_r} \right\}.$$

A.3. Monte Carlo results

Table 5. Detailed simulation results. RMSE $= \sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\pi}_r - \pi)^2}$, MAE $= R^{-1} \sum_{r=1}^{R} |\hat{\pi}_r - \pi|$, where $\hat{\pi}_r$ denotes the coefficient estimate obtained in replication r and π the true parameter. l.b. denotes the parameter lower bound (1×10^{-5}).

	Exponential-ACD			Weibull-ACD				Burr-ACD				
	ω	α	β	ω	α	β	γ	ω	α	β	κ	σ^2
DGPI												
0.10 Qtl.	0.1728	0.0889	0.6643	0.1729	0.0890	0.6644	0.9922	0.1724	0.0889	0.6648	0.9937	l.b.
0.50 Qtl.	0.1999	0.0994	0.6998	0.2000	0.0994	0.6999	1.0002	0.1993	0.0993	0.7004	1.0024	l.b.
0.90 Qtl.	0.2307	0.1095	0.7341	0.2307	0.1096	0.7340	1.0086	0.2304	0.1094	0.7342	1.0132	0.0144
Mean	0.2007	0.0996	0.6997	0.2007	0.0996	0.6997	1.0003	0.2004	0.0996	0.6999	1.0031	0.0044
Std. Dev.	0.0221	0.0078	0.0267	0.0221	0.0078	0.0267	0.0063	0.0222	0.0078	0.0270	0.0076	0.0070
RMSE	0.0221	0.0078	0.0267	0.0221	0.0078	0.0267	0.0063	0.0222	0.0078	0.0270	0.0083	0.0083
MAE	0.0174	0.0061	0.0212	0.0174	0.0061	0.0212	0.0050	0.0175	0.0061	0.0213	0.0064	0.0044
DGP II												
0.10 Qtl.	0.0079	0.0170	0.9641	0.0079	0.0170	0.9641	0.9922	0.0079	0.0170	0.9643	0.9937	l.b.
0.50 Qtl.	0.0104	0.0199	0.9696	0.0104	0.0199	0.9696	1.0002	0.0103	0.0199	0.9696	1.0025	l.b.
0.90 Qtl.	0.0138	0.0230	0.9743	0.0138	0.0230	0.9743	1.0086	0.0138	0.0229	0.9743	1.0133	0.0145
Mean	0.0107	0.0199	0.9694	0.0107	0.0199	0.9694	1.0003	0.0107	0.0199	0.9694	1.0031	0.0044
Std. Dev.	0.0025	0.0023	0.0041	0.0025	0.0023	0.0041	0.0063	0.0025	0.0023	0.0041	0.0077	0.0070
RMSE	0.0026	0.0023	0.0041	0.0026	0.0023	0.0041	0.0063	0.0026	0.0023	0.0041	0.0083	0.0082
MAE	0.0019	0.0018	0.0032	0.0019	0.0018	0.0032	0.0050	0.0019	0.0018	0.0032	0.0064	0.0044
DGP III												
0.10 Qtl.	0.1707	0.0869	0.6580	0.1716	0.0872	0.6613	0.5953	0.1716	0.0874	0.6554	0.5962	l.b.
0.50 Qtl.	0.1995	0.0994	0.6995	0.1995	0.0994	0.6997	0.6001	0.2003	0.0996	0.6993	0.6015	l.b.
0.90 Qtl.	0.2341	0.1121	0.7385	0.2310	0.1118	0.7372	0.6051	0.2368	0.1127	0.7370	0.6079	0.0143
Mean	0.2008	0.0997	0.6993	0.2010	0.0997	0.6991	0.6002	0.2025	0.1000	0.6970	0.6018	0.0044
Std. Dev.	0.0252	0.0100	0.0312	0.0241	0.0095	0.0299	0.0038	0.0256	0.0097	0.0326	0.0046	0.0070
RMSE	0.0252	0.0100	0.0312	0.0241	0.0095	0.0299	0.0038	0.0257	0.0097	0.0328	0.0049	0.0082
MAE	0.0199	0.0079	0.0247	0.0191	0.0075	0.0237	0.0030	0.0203	0.0077	0.0256	0.0038	0.0044
DGP IV												
0.10 Qtl.	0.0078	0.0169	0.9639	0.0081	0.0171	0.9637	0.5954	0.0081	0.0171	0.9638	0.5962	l.b.
0.50 Qtl.	0.0104	0.0199	0.9696	0.0106	0.0200	0.9693	0.6001	0.0105	0.0200	0.9693	0.6013	l.b.
0.90 Qtl.	0.0139	0.0233	0.9742	0.0138	0.0231	0.9740	0.6052	0.0138	0.0231	0.9740	0.6078	0.0140
Mean	0.0108	0.0200	0.9692	0.0108	0.0201	0.9690	0.6002	0.0108	0.0201	0.9690	0.6018	0.0042
Std. Dev.	0.0026	0.0024	0.0043	0.0024	0.0023	0.0041	0.0038	0.0024	0.0024	0.0041	0.0046	0.0068
RMSE	0.0027	0.0024	0.0044	0.0026	0.0024	0.0042	0.0038	0.0026	0.0024	0.0042	0.0049	0.0079
MAE	0.0020	0.0020	0.0034	0.0019	0.0019	0.0032	0.0030	0.0019	0.0019	0.0032	0.0038	0.0042
DGP V												
Mean	0.2024	0.1099	0.6820	0.1679	0.0840	0.6973	0.7490	0.2006	0.1000	0.6988	1.9976	1.4956
Std. Dev.	0.0896	0.0893	0.1201	0.0294	0.0186	0.0481	0.0235	0.0145	0.0074	0.0166	0.0266	0.0380
RMSE	0.0896	0.0898	0.1214	0.0436	0.0245	0.0481	1.2512	0.0145	0.0074	0.0166	0.0267	0.0382
MAE	0.0605	0.0440	0.0823	0.0363	0.0207	0.0371	1.2510	0.0115	0.0060	0.0134	0.0212	0.0305
DGP VI												
Mean	0.0138	0.0261	0.9607	0.0089	0.0169	0.9688	0.7490	0.0104	0.0205	0.9687	1.9897	1.4850
Std. Dev.	0.0242	0.0526	0.0544	0.0019	0.0034	0.0052	0.0235	0.0009	0.0015	0.0018	0.0264	0.0370
RMSE	0.0245	0.0529	0.0552	0.0022	0.0046	0.0053	1.2512	0.0010	0.0016	0.0022	0.0283	0.0400
MAE	0.0066	0.0121	0.0148	0.0018	0.0039	0.0041	1.2510	0.0008	0.0013	0.0017	0.0229	0.0323