Non-parametric specification tests for conditional duration models

Marcelo Fernandes
European University Institute
mfernand@iue.it

Joachim Grammig
University of Frankfurt
grammig@wiwi.uni-frankfurt.de

First Draft: January 2000
Current Draft: February 2000

Abstract. This paper deals with the estimation and testing of conditional duration models by looking at the density and baseline hazard rate functions. More precisely, we focus on the distance between the parametric density (or hazard rate) function implied by the duration process and its non-parametric estimate. Asymptotic justification is derived using the functional delta method for fixed and gamma kernels, whereas finite sample properties are investigated through Monte Carlo simulations. Finally, we show the practical usefulness of such testing procedures by carrying out an empirical assessment of whether autoregressive conditional duration models are appropriate tools for modelling price durations of stocks traded at the New York Stock Exchange.

JEL Classification: C14, C51, C52.
Keywords: duration, functional delta method, gamma kernel, hazard rate.

Acknowledgements. We are indebted to Søren Johansen and seminar participants at Universidad Carlos III de Madrid for valuable comments. We also acknowledge with gratitude financial support from CNPq-Brazil (grant 200608/95-9) and the German Research Foundation, the BMW foundation and Roland Berger & Partners, respectively. The usual disclaimer applies.

1 Introduction

The availability of financial transactions data hoisted the interest in applied microstructure research. Thinning raw data enables analysts to define the events of interest, e.g. quote updates and limit-order execution, and then compute the corresponding waiting times. Typically, the resulting duration processes are influenced by public and private information, what motivates the use of conditional duration models. Therefore, it is not surprising that microstructure studies employing conditional duration models abound in the literature (e.g. Engle and Lange, 1997; Lo, MacKinlay and Zhang, 1997; Lunde, 1999). In particular, price durations are closely linked to the instantaneous volatility of the mid-quote price process (Engle and Russell, 1997). Besides, price durations play an interesting role in option pricing as well (Pringent, Renault and Scaillet, 1999). Trade and volume durations mirror in turn features such as market liquidity and the information arrival rate (Gouriéroux, Jasiak and Le Fol, 1996).

Engle and Russell's (1998) autoregressive conditional duration (ACD) model is the starting point of such analyses, though there are several extensions. Engle (1996) and Ghysels and Jasiak (1998a) combine conditional duration models with GARCH-type effects, whereas Ghysels, Gouriéroux and Jasiak (1997) introduce a stochastic volatility duration model to cope with higher order dynamics in the duration process. Ghysels and Jasiak (1998b) investigate the persistence of intra-trade durations using a fractionally integrated ACD model, whilst Zhang, Russell and Tsay (1999) advocate for a non-linear version of the ACD model rooted in a self exciting threshold autoregressive framework. Bauwens and Veredas (1999), Grammig and Maurer (1999), Lunde (1999b), and Hamilton and Jorda (1999) argue for conditional duration models that accommodate more flexible hazard rate functions. Bauwens and Giot's (1997) logarithmic ACD model provides a more suitable framework for testing market microstructure hypotheses as it avoids some of the parameter restrictions implied by the original ACD specification. 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.

Despite the recent boom of empirical applications, the literature has devoted so far little attention to testing the specification of conditional duration models. The practice is to perform simple diagnostic tests to check whether the standardised residuals are independent and identically distributed (iid). If, on the one hand, all papers use the Ljung-Box statistic to test for serial correlation; on the other hand, only a few tests whether the distribution of the error term is correctly specified. Engle and Russell (1998) and Grammig, Hujer, Kokot and Maurer (1998) check the first and second moments of the residuals with particular attention to measuring excess dispersion, whilst others use QQ-plots (Bauwens and Veredas, 1999) and Bartlett identity tests (Pringent et al., 1999). Grammig and Wellner (1999) take a different approach by estimating and testing conditional duration models using a GMM framework. More recently, Bauwens, Giot, Grammig and Veredas (2000) employ the techniques developed by Diebold, Gunther and Tay (1998) to evaluate density forecasts.

Misspecification of the distribution of the error process may seem unimportant given that quasi maximum likelihood (QML) methods provide consistent estimates (Engle, 1996). However, QML estimation of conditional duration models may perform quite poorly in finite samples. Consider, for instance, a model in which standardised durations have a distribution that engenders a non-monotonic baseline hazard rate function. Quasi maximum likelihood methods rooted in distributions with monotonic hazard rates will then fail to produce sound estimates even in quite large samples such as 15000 observations (Grammig and Maurer, 1999). The poor performance of QML estimation has quite serious implications for models that attempt to uncover the link between duration and volatility, e.g. Ghysels and Jasiak's (1998) ACD-GARCH process. Indeed, shoddy estimates of the expected duration may produce rather misleading results for the volatility process.

This paper develops tools to test the distribution of the error term in a conditional duration model. We propose testing procedures that gauge the closeness between non- and parametric estimates of the density and baseline hazard rate functions of the standardised durations. There is no novelty in the idea of comparing a consistent estimator under correct parameterisation to another which is consistent even if the model is misspecified. It constitutes, for instance, the hinge of Hausman's (1978) specification tests and Aït-Sahalia's (1996) density matching approach to estimate and test diffusion processes.

Our tests carry some interesting properties. In contrast to Bartlett identity tests (Chesher, Dhaene, Gouriéroux and Scaillet, 1999), it examines the whole distribution of the standardised residuals instead of a small number of moment restrictions. In addition, our tests are nuisance parameter free in that there is no asymptotic cost in replacing errors with estimated residuals. Further, as all results are derived under mixing conditions, there is no need to carry out a previous test for serial independence of the standardised errors. This is quite convenient in view that a joint test such as the GMM overidentification test does not pinpoint the cause of rejection. Lastly, Monte Carlo simulations indicate that some versions of our tests are quite promising in terms of finite sample size and power.

The remainder of this paper is organised as follows. Section 2 describes the family of conditional duration models we have in mind. Section 3 discusses the design of the testing procedures. Section 4 deals with the limiting behaviour of such tests. First, we show asymptotic normality under the null hypothesis that the conditional duration model is properly specified. Second, we compute the asymptotic local power by considering a sequence of local alternatives. Third, we derive the conditions in which our tests are nuisance parameter free. Section 5 investigates finite sample properties through Monte Carlo simulations. Section 6 tests whether ACD models are suitable to model price durations of frequently traded stocks at the New York Stock Exchange (NYSE). In section 7, we summarise the results and offer concluding remarks. For ease of exposition, an appendix collects all proofs and technical lemmas.

2 Conditional duration models

Let $x_i = \psi_i \epsilon_i$, where the duration $x_i = t_i - t_{i-1}$ denotes the time elapsed between events occurring at time t_i and t_{i-1}, the conditional duration process $\psi_i \propto E\big(x_i \,|\, I_{i-1}\big)$ is independent of ϵ_i and I_{i-1} is the set including all information available at time t_{i-1}. To nest the existing ACD models, we consider the following general specification for the conditional expectation

$$\psi_i = g(\psi_{i-1}, \epsilon_{i-1}, u_i; \phi), \tag{1}$$

where $u_i | I_{i-1} \sim N(0, \sigma_u^2)$ and ϕ is a vector of parameters. If the interest rests on modelling microstructure, one may incorporate additional predetermined variables as well (Bauwens and Giot, 1997 and 1998; Engle and Russell, 1998).

Further, suppose that ϵ_i is iid with Burr density

$$f_B\left(\epsilon_i,\, \theta_B\right) = \frac{\kappa\, \xi_B^{\kappa}\, \epsilon_i^{\kappa-1}}{\left(1+\sigma^2\, \xi_B^{\kappa}\, \epsilon_i^{\kappa}\right)^{1+1/\sigma^2}}, \tag{2}$$

with $\kappa > \sigma^2 > 0$ and mean

$$\xi_B \equiv \frac{\Gamma\big(1+1/\kappa\big)\, \Gamma\big(1/\sigma^2 - 1/\kappa\big)}{\sigma^{2(1+1/\kappa)}\, \Gamma\big(1+1/\sigma^2\big)}.$$

It is readily seen that the conditional density of x_i is also Burr with parameter vector $\left(\xi_B^{\kappa}\psi_i^{-\kappa}, \kappa, \sigma^2\right)$. Accordingly, the conditional hazard rate function reads

$$\Gamma_B\left(x_i \,\middle|\, I_{i-1};\, \theta_B\right) = \frac{\kappa\, \xi_B^{\kappa}\, \psi_i^{-\kappa}\, x_i^{\kappa-1}}{1+\sigma^2\, \xi_B^{\kappa}\, \psi_i^{-\kappa}\, x_i^{\kappa}}, \tag{3}$$

which is non-monotonic with respect to the standardised duration if $\kappa > 1$.

When σ^2 shrinks to zero, (2) reduces to a Weibull distribution, viz.

$$f_W\left(\epsilon_i,\, \theta_W\right) = \kappa\, \xi_W^{\kappa}\, \epsilon_i^{\kappa-1} \exp\left(-\xi_W^{\kappa}\, \epsilon_i^{\kappa}\right),$$

where $\xi_W = \Gamma(1+1/\kappa)$. Accordingly, the conditional distribution of the duration process is also Weibull and the conditional hazard rate function reads $\Gamma_W\left(x_i \,\middle|\, I_{i-1};\, \theta_W\right) = \kappa\xi_W^{\kappa}\, \psi_i^{-\kappa} x_i^{\kappa-1}$. In contrast to the Burr case, the conditional hazard rate implied by the Weibull distribution is monotonic. It decreases with the standardised duration for $0 < \kappa < 1$, increases for $\kappa > 1$ and remains constant for $\kappa = 1$. In the latter case, the Weibull coincide with the exponential distribution and the conditional hazard rate function of the duration process is simply $\Gamma_E\left(x_i \,\middle|\, I_{i-1};\, \theta_E\right) = \psi_i^{-1}$. Albeit Engle and Russell (1998) suggest the use of exponential and Weibull distributions, the Burr ACD model seems to deliver better results for price durations (Bauwens et al., 2000).

3 Specification tests

As conditional duration models are usually estimated by QML methods, likelihood ratio tests are available to compare nested distributions in conditional duration models. However, due to the presence of inequality constraints in the parameter space, the limiting distribution of the test statistic is a mixing of χ^2-distributions with probability weights depending on the variance of the parameter estimates (Wolak, 1991). Accordingly, it is extremely difficult to obtain empirically implementable asymptotically exact critical values. As an alternative, Wolak (1991) suggests applying asymptotic bounds tests, but bounds are in most instances quite slack, yielding inconclusive results more likely.

In the following, we design a simple testing strategy which checks specification by matching density functionals. More precisely, we test the null

$$H_0: \quad \exists\, \theta_0 \in \Theta \;\text{ such that }\; f(\cdot, \theta_0) = f(\cdot) \tag{4}$$

against the alternative hypothesis that there is no such $\theta_0 \in \Theta$. The true density $f(\cdot)$ of the standardised durations is of course unknown, otherwise we could merely check whether it belongs to the proposed parametric family of distributions. Accordingly, we 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. It is therefore natural to carry a test by gauging the closeness between these two density estimates.

For that purpose, we consider the distance

$$\Psi_f = \int_0^\infty \mathbb{1}(x \in \mathcal{S}) \big[f(x, \theta) - f(x)\big]^2 f(x)\, \mathrm{d}x \tag{5}$$

to build a first testing procedure, which we label the D-test. We introduce the compact subset $\mathcal{S}$ to avoid regions in which density estimation is unstable. The sample analog reads

$$\Psi_{\hat{f}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \in \mathcal{S}) \left[f\big(x_i, \hat{\theta}\big) - \hat{f}(x_i)\right]^2, \tag{6}$$

where $\hat{\theta}$ and $\hat{f}(\cdot)$ denote consistent estimates of the true parameter θ_0 and density $f(\cdot)$, respectively. The null hypothesis is then rejected if the D-test statistic $\Psi_{\hat{f}}$ is large enough.

By virtue of the one-to-one mapping linking hazard rate and density functions, the null hypothesis (4) implies that there exists $\theta_0 \in \Theta$ such that the hazard rate function implied by the parametric model $\Gamma_{\theta_0}(\cdot)$ equals the true hazard function $\Gamma_f(\cdot)$. Accordingly, we consider a second test based on the statistic

$$\Lambda_{\hat{f}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x_i \in \mathcal{S}) \left[\Gamma_{\hat{\theta}}(x_i) - \Gamma_{\hat{f}}(x_i)\right]^2, \tag{7}$$

which we refer as the H-test. To provide a minimum-distance flavour to both D- and H-tests, one may estimate the parametric model by minimising (6) and

(7), respectively. Though we derive in the next section the limiting behaviour of the resulting M-estimators $\hat{\theta}_n^D = \text{argmin}_{\theta\in\Theta}\Psi_{\hat{f}}$ and $\hat{\theta}_n^H = \text{argmin}_{\theta\in\Theta}\Lambda_{\hat{f}}$, we rather avoid tackling identification issues to keep focus on testing.

4 Asymptotic justification

In what follows, we derive asymptotic results for the test statistics and their implied M-estimators using Aït-Sahalia's (1994) functional delta method. In fact, the limiting behaviour of the D-test was originally developed by Bickel and Rosenblatt (1973), who assume random sampling. Aït-Sahalia (1996) extends Bickel and Rosenblatt's results to mixing processes to build a specification test for diffusion processes, and shows the asymptotic normality of the implied M-estimator. Accordingly, the set of assumptions we impose is quite similar and the asymptotics are the same up to a weighting scheme. Before moving to the details of the asymptotic theory, it is noteworthy that the M-estimators implied by the D- and H-tests hinge on a two-step procedure in which the first step involves a kernel estimation and the second step solves a minimisation problem. As such, these estimators belong to the class of M-estimators discussed in Newey (1994).

4.1 Assumptions

Consider a real-valued random variable x_i with discretely sampled observations $x_1, \ldots, x_n$. We consider the following set of regularity conditions.

A1 The sequence $\{x_i\}$ is strictly stationary and β-mixing with $\beta_j = O(j^{-\delta})$, where $\delta > 1$. Further, $E||x_i||^k < \infty$ for some constant $k > 2\delta/(\delta - 1)$.

A2 The density function $f_x = f(x)$ of x_i is continuously differentiable up to order $s+1$ and its derivatives are bounded and square-integrable. Further, f_x is bounded away from zero on the compact interval $\mathcal{S}$, i.e. $\inf_{\mathcal{S}} f_x > 0$.

A3 The fixed kernel K is of order s (even integer) and is continuously differentiable up to order s on $\mathbb{R}$ with derivatives in $L^2(\mathbb{R})$. Let $e_K \equiv \int_u K^2(u)\mathrm{d}u$ and $v_K \equiv \int_v \left[\int_u K(u)K(u+v)\mathrm{d}u\right]^2 \mathrm{d}v$.

A4 As the sample size n grows, the bandwidths for the fixed and gamma kernels are such that $h_n = o\left(n^{-2/(4s+1)}\right)$ and $b_n = o\left(n^{-4/9}\right)$, respectively.

A5 The parameter space $\Theta \subset \mathbb{R}^k$ is compact. Let $\zeta(\cdot, \theta)$ denote the density function $f(\cdot, \theta)$ for the D-test and the baseline hazard rate function $\Gamma(\cdot, \theta)$ for the H-test. In a neighbourhood of the true parameter θ_0, $\zeta(\cdot, \theta)$ is twice continuously differentiable in θ, the matrix $E\left[\frac{\partial}{\partial \theta}\zeta(\cdot, \theta)\frac{\partial}{\partial \theta'}\zeta(\cdot, \theta)\right]$ has full rank, and $\frac{\partial^2}{\partial \theta_i \partial \theta_j}\zeta(\cdot, \theta)$ is bounded in absolute value for every i, j and $\theta \in \Theta$.

A6 Consider f_* and f_+ in a neighbourhood N_f of the true density f_x. Then, the leading term ϑ_f that drives the asymptotic distribution of the implied M-estimators is such that

$$
\begin{array}{ll}
(i) & E\left|\vartheta_f\right|^{3+r} < \infty, \text{ for } r > (3+\eta)(3+\eta/2)/\eta, \ \forall \eta > 0 \\
(ii) & E \sup\limits_{f_* \in N_f} \left|\vartheta_{f_*}\right|^2 < \infty \\
(iii) & E\left|\vartheta_{f_*} - \vartheta_{f_+}\right|^2 \leq c\|f_* - f_+\|^2_{L(\infty,m)},
\end{array}
$$

where c is a constant, $\|\cdot\|_{L(\infty,m)}$ denotes the Sobolev norm of order (∞, m) and m is an integer such that $0 < m < s/2 + 1/4$.

Assumption A1 restricts the amount of dependence allowed in the observed data sequence in order to ensure that the central limit theorem holds. As usual, there is a trade-off between the number of existing moments and the admissible level of dependence. Carrasco and Chen (1999) offer more details concerning the β-mixing properties of ACD models. Assumption A2 requires that the density function is smooth enough to admit a functional Taylor expansion. Though assumption A3 provides enough room for higher order kernels, in what follows, we implicit assume that the kernel is of second order (i.e. $s = 2$). Assumption A4 induces some degree of undersmoothing to force the asymptotic biases of the test statistics to vanish. Further, it implies that the gamma kernel bandwidth b_n is of the same order of h_n^2 for second order kernels (see Chen, 2000). Assumptions A5 ensures that the M-estimators θ_f^D and θ_f^H are well defined. Finally, A6 guarantees that one can estimate consistently the asymptotic variance of the M-estimators using a non-parametric correction à la Newey and West (1987).

4.2 Matching the density function

The D-test gauges the discrepancy between the parametric and non-parametric estimates of the stationary density. The functional of interest is

$$\Psi_{\hat{f}} = \int_x \mathbb{1}(x \in \mathcal{S}) \left[f(x, \theta_{\hat{f}}) - \hat{f}(x) \right]^2 \hat{f}(x) \, \mathrm{d}x, \tag{8}$$

where $\mathbb{1}(\cdot)$ is the indicator function and θ_f is the functional implied by the estimator of θ. Assume further that it admits the following functional expansion

$$\Psi_{\hat{f}} = \Psi_f + \mathrm{D}\Psi_f(h_x) + \frac{1}{2}\mathrm{D}^2\Psi_f(h_x, h_x) + O\left(\|h_x\|^3\right), \tag{9}$$

where $h_x = \hat{f}_x - f_x$ and $\|\cdot\|$ denotes the L^2 norm. By the Riesz representation theorem, the functional derivative $\mathrm{D}\Psi_f(\cdot)$ has a dual representation of the form $\mathrm{D}\Psi_f(h_x) = \int_x \psi_f(x) h_x \, \mathrm{d}x$. It follows from Aït-Sahalia's (1994) functional delta method that ψ_f stands for the leading term that drives the asymptotic distribution of $\Psi_{\hat{f}}$. If the first functional derivative is degenerate, then the asymptotic distribution is driven by the second order term of the expansion.

Let f_x and $f_{x,\theta}$ denote the true and parametric density functions, respectively. The first functional derivative of Ψ_f reads

$$\mathrm{D}\Psi_f(h_x) = \int_{\mathcal{S}} (f_{x,\theta} - f_x)^2 h_x \, \mathrm{d}x + 2 \int_{\mathcal{S}} \left[\frac{\partial f_{x,\theta}}{\partial \theta} \mathrm{D}\theta_f(h_x) - h_x \right] (f_{x,\theta} - f_x) f_x \, \mathrm{d}x,$$

where $\mathrm{D}\theta_f(\cdot)$ denotes the first derivative of the functional θ_f implied by the estimator under consideration. As $\mathrm{D}\Psi_f(h_x)$ is singular under the null, the limiting distribution of $\Psi_{\hat{f}}$ depends on the second functional derivative, namely

$$\begin{aligned} \mathrm{D}^2\Psi_f(h_x, h_x) \quad = \quad & 2 \int_{\mathcal{S}} \frac{\partial f(x,\theta_f)}{\partial \theta} \frac{\partial f(x,\theta_f)}{\partial \theta'} \left[\mathrm{D}\theta_f(h_x) \right]^2 f_x \, \mathrm{d}x \\ & - 4 \int_{\mathcal{S}} \frac{\partial f(x,\theta_f)}{\partial \theta} \mathrm{D}\theta_f(h_x) \, f_x h_x \, \mathrm{d}x + 2 \int_{\mathcal{S}} f_x h_x^2 \, \mathrm{d}x. \end{aligned} \tag{10}$$

However, the first and second terms of the right-hand side do not play a role in the asymptotic distribution of the test statistic. The functional delta method shows indeed that the asymptotics is driven by the unsmoothest term of the first non-degenerate derivative for it converges at a slower rate. The third term contains a Dirac mass in its inner product representation, and thus will lead the asymptotics.

Theorem 1. *Under the null and assumptions A1 to A4, the statistic*

$$\hat{\tau}_n^D = \frac{n h_n^{1/2} \Psi_{\hat{f}} - h_n^{-1/2} \hat{\delta}_D}{\hat{\sigma}_D} \xrightarrow{d} N(0,1),$$

where $\hat{\delta}_D$ *and* $\hat{\sigma}_D^2$ *are consistent estimates of* $\delta_D = e_K\, E\big[\mathbb{1}(x \in \mathcal{S}) f_x\big]$ *and* $\sigma_D^2 = v_K\, E\big[\mathbb{1}(x \in \mathcal{S}) f_x^3\big]$*, respectively.*

Proof. See Aït-Sahalia (1996).

As the time elapsed between transactions is strictly positive, durations have a support which is bounded from below. Further, the bulk of duration data is typically in the vicinity of the origin. Accordingly, $\hat{\tau}_n^D$ 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 and density is easily derived: indeed, if $Y = \log X$, then $f_Y(y) = f_X\big[\exp(y)\big]\exp(y)$. Alternatively, one may utilise asymmetric kernels to benefit from the fact that they never assign weight outside the density support (Chen, 2000). In particular, the gamma kernel

$$K_{x/b_n+1,b_n}(u) = \frac{u^{x/b_n}\exp(-u/b_n)}{\Gamma(x/b_n+1)b_n^{x/b_n}}\mathbb{1}\big\{u \in [0,\infty)\big\} \tag{11}$$

with bandwidth b_n is quite convenient to handle a density function whose support is bounded from the origin. Therefore, we consider a second version of the D-test in which the density estimation uses a gamma kernel.

Theorem 2. *Under the null and assumptions A1 to A4, the statistic*

$$\tilde{\tau}_n^D = \frac{n b_n^{1/4}\Psi_{\tilde{f}} - b_n^{-1/4}\tilde{\delta}_G}{\tilde{\sigma}_G} \xrightarrow{d} N(0,1),$$

where $\tilde{\delta}_G$ *and* $\tilde{\sigma}_G^2$ *are consistent estimates of* $\delta_G = \frac{1}{2\sqrt{\pi}}E\big[\mathbb{1}(x \in \mathcal{S})x^{-1/2}f_x\big]$ *and* $\sigma_G^2 = \frac{1}{\sqrt{\pi}}E\big[\mathbb{1}(x \in \mathcal{S})x^{-1/2}f_x^3\big]$*, respectively.*

Consider now the following sequence of local alternatives

$$H_{1n}^D: \quad \sup_{x\in\mathcal{S}}\left|f^{[n]}(x,\theta) - f^{[n]}(x) - \varepsilon_n \ell_D(x)\right| = o(\varepsilon_n), \tag{12}$$

where $\|f^{[n]} - f\| = o\left(n^{-1}h_n^{-1/2}\right)$, $\varepsilon_n = n^{-1/2}h_n^{-1/4}$ and $\ell_D(x)$ is such that $\ell_D^{\mathcal{S}} \equiv E\big[\mathbb{1}(x \in \mathcal{S})\ell_D^2(x)\big]$ exists and $E\big[\ell_D(x)\big] = 0$. The next result illustrates the fact that both versions of the D-test have non-trivial power under local alternatives that shrink to the null at rate ε_n.

Theorem 3. *Under the sequence of local alternatives* H_{1n}^D *and assumptions A1 to A4,* $\hat{\tau}_n^D \xrightarrow{d} N\left(\ell_D^{\mathcal{S}}/\sigma_D, 1\right)$*, whereas* $\tilde{\tau}_n^D \xrightarrow{d} N\left(\ell_D^{\mathcal{S}}/\sigma_G, 1\right)$.

To maximise power of both versions of the D-test, one could consider the most favourable scenario to the parametric model by utilising the M-estimator $\hat{\theta}_n^D$. The corresponding implicit functional is then

$$\int_{\mathcal{S}} \frac{\partial f(x, \theta_f^D)}{\partial \theta} \left[f\left(x, \theta_f^D\right) - f(x) \right] f(x)\, \mathrm{d}x \equiv 0, \tag{13}$$

which produces

$$\mathrm{D}\theta_f^D(h_x) = \left\{ \int_{\mathcal{S}} \frac{\partial f(x,\theta)}{\partial \theta} \frac{\partial f(x,\theta)}{\partial \theta'} f(x) \mathrm{d}x \right\}^{-1} \int_{\mathcal{S}} \frac{\partial f(x,\theta)}{\partial \theta} f(x) h(x) \mathrm{d}x. \tag{14}$$

Accordingly, the limiting distribution is driven by

$$\vartheta_f^D(x) = \mathbb{1}(x \in \mathcal{S}) \left\{ \int_{\mathcal{S}} \frac{\partial f(x,\theta)}{\partial \theta} \frac{\partial f(x,\theta)}{\partial \theta'} f(x)\, \mathrm{d}x \right\}^{-1} \frac{\partial f(x,\theta)}{\partial \theta} f(x). \tag{15}$$

Theorem 4. *Under the null and assumptions A1 to A5, $n^{1/2}\big(\hat{\theta}_f^D - \theta_0\big) \xrightarrow{d} N(0, \Omega_D)$, where $\Omega_D = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left[\vartheta_f^D(x_i), \vartheta_f^D(x_{i+k})\right]$ is the long run covariance matrix of ϑ_f^D. In addition, if assumption A6 holds, it suffices to plug $\hat{\theta}_f^D$ into ϑ_f^D and truncate the infinite sum as in Newey and West (1987) to obtain a consistent estimator of the asymptotic variance.*

Proof. See Aït-Sahalia (1996).

4.3 Matching the baseline hazard rate function

The H-test compares the parametric and non-parametric estimates of the baseline hazard rate. The motivation is simple. The usual densities associated with duration models, e.g. exponential, Weibull and Burr, may engender fairly similar shapes depending on the parameter values. In turn, they hatch very different hazard rate functions: it is flat for the exponential, monotonic for the Weibull and non-monotonic for the Burr.

The functional of interest reads

$$\Lambda_{\hat{f}} = \int_{\mathcal{S}} \left[\Gamma_\theta(x) - \Gamma_{\hat{f}}(x) \right]^2 \hat{f}_x \, \mathrm{d}x, \tag{16}$$

Suppose that (16) admits a second order Taylor expansion about the true density, viz.

$$\Lambda_{\hat{f}} = \Lambda_f + \mathrm{D}\Lambda_f(h_x) + \frac{1}{2}\mathrm{D}^2\Lambda_f(h_x, h_x) + O\left(\|h_x\|^3\right), \tag{17}$$

where $\Lambda_f = \int_S \left[\Gamma_\theta(x) - \Gamma_f(x)\right]^2 f_x \,\mathrm{d}x$ and $h_x = \hat{f}_x - f_x$ as before. The first functional derivative is then

$$\begin{aligned} \mathrm{D}\Lambda_f(h_x) &= \int_S \left[\Gamma_\theta(x) - \Gamma_f(x)\right]^2 h_x \,\mathrm{d}x \\ &\quad + 2\int_S \left[\Gamma_\theta(x) - \Gamma_f(x)\right]\left[\frac{\partial \Gamma_\theta(x)}{\partial \theta}\mathrm{D}\theta_f(h_x) - \mathrm{D}\Gamma_f(h_x)\right] f_x \,\mathrm{d}x, \quad (18) \end{aligned}$$

where

$$\mathrm{D}\Gamma_f(h_x) = \frac{h(x) - \Gamma_f(x)\int_x 1\!\!1(u < x)h(u)\mathrm{d}u}{S_x} \tag{19}$$

and S_x denotes the survival function $1 - F(x)$. It is readily seen that, if the baseline hazard is properly specified, the first derivative is singular.

The asymptotic distribution of the H-test relies then on the second order functional derivative, which under the null reads

$$\begin{aligned} \mathrm{D}^2\Lambda_f(h_x, h_x) &= 2\int_S \left[\mathrm{D}\Gamma_f(h_x)\right]^2 f_x \,\mathrm{d}x \\ &\quad + 2\int_S \frac{\partial \Gamma_\theta(x)}{\partial \theta}\frac{\partial \Gamma_\theta(x)}{\partial \theta'}\left[\mathrm{D}\theta_f(h_x)\right]^2 f_x \,\mathrm{d}x \\ &\quad - 4\int_S \frac{\partial \Gamma_\theta(x)}{\partial \theta}\mathrm{D}\theta_f(h_x)\mathrm{D}\Gamma_f(h_x) f_x \,\mathrm{d}x. \quad (20) \end{aligned}$$

It turns out that the first term leads the asymptotics as it contains the unsmoothest term of the expansion.

Theorem 5. *Under the null and assumptions A1 to A4, the statistic*

$$\hat{\tau}_n^H = \frac{n h_n^{1/2}\Lambda_{\hat{f}} - h_n^{-1/2}\hat{\lambda}_H}{\hat{\varsigma}_H} \xrightarrow{d} N(0,1),$$

where $\hat{\lambda}_H$ and $\hat{\varsigma}_H^2$ are consistent estimates of $\lambda_H = e_K\, E\big[1\!\!1(x \in \mathcal{S})\Gamma_f(x)/S_x\big]$ and $\varsigma_H^2 = v_K\, E\big[1\!\!1(x \in \mathcal{S})\Gamma_f^3(x)/S_x\big]$, respectively.

In contrast to the density function, in general, there is no closed form solution for the hazard rate of the log-standardised duration. One may of course solve it by numerical integration, though at the expense of simplicity. Notwithstanding, it is straightforward to fashion the H-test to gamma kernels.

Theorem 6. *Under the null and assumptions A1 to A4, the statistic*

$$\tilde{\tau}_n^H = \frac{n b_n^{1/4}\Lambda_{\tilde{f}} - b_n^{-1/4}\tilde{\lambda}_G}{\tilde{\varsigma}_G} \xrightarrow{d} N(0,1),$$

where $\tilde{\lambda}_G$ and $\tilde{\varsigma}_G^2$ estimate consistently $\lambda_G = \frac{1}{2\sqrt{\pi}}E\left[1\!\!1(x \in \mathcal{S})x^{-1/2}\Gamma_f(x)/S_x\right]$ and $\varsigma_G^2 = \frac{1}{\sqrt{\pi}}E\left[1\!\!1(x \in \mathcal{S})x^{-1/2}\Gamma_f^3(x)/S_x\right]$, respectively.

Consider next the following sequence of local alternatives

$$H_{1n}^{H}: \quad \sup_{x\in\mathcal{S}} \left|\Gamma^{[n]}(x,\theta) - \Gamma_f^{[n]}(x) - \varepsilon_n \ell_H(x)\right| = o(\varepsilon_n), \tag{21}$$

where $\left\|\Gamma_f^{[n]} - \Gamma_f\right\| = o\left(n^{-1} h_n^{-1/2}\right)$, $\varepsilon_n = n^{-1/2} h_n^{-1/4}$ and $\ell_H(x)$ is such that $\ell_H^{\mathcal{S}} \equiv E\left[\mathbb{1}(x \in \mathcal{S})\ell_H^2(x)\right] < \infty$ and $E\left[\ell_H(x)\right] = 0$. It follows then that both versions of the H-test can distinguish alternatives that get closer to the null at rate ε_n while maintaining constant power level.

Theorem 7. *Under the sequence of local alternatives H_{1n}^{H} and assumptions A1 to A4, $\hat{\tau}_n^H \xrightarrow{d} N\left(\ell_H^{\mathcal{S}}/\varsigma_H, 1\right)$, whereas $\tilde{\tau}_n^H \xrightarrow{d} N\left(\ell_H^{\mathcal{S}}/\varsigma_G, 1\right)$.*

Finally, consider the M-estimator θ_f^H that minimises the distance between the non- and parametric estimates of the baseline hazard rate function. The corresponding implicit functional is

$$\int_{\mathcal{S}} \frac{\partial\Gamma(x,\theta_f^H)}{\partial\theta}\left[\Gamma(x,\theta_f^H) - \Gamma_f(x)\right] f(x)\,\mathrm{d}x \equiv 0, \tag{22}$$

which results in the following first derivative

$$\mathrm{D}\theta_f^H(h_x) = \left\{\int_{\mathcal{S}} \frac{\partial\Gamma(x,\theta)}{\partial\theta}\frac{\partial\Gamma(x,\theta)}{\partial\theta'} f(x)\,\mathrm{d}x\right\}^{-1} \int_{\mathcal{S}} \frac{\partial\Gamma_\theta(x)}{\partial\theta}\,\mathrm{D}\Gamma_f(h_x) f(x)\,\mathrm{d}x. \tag{23}$$

From (19), it is readily seen that

$$\vartheta_f^H(x) = \mathbb{1}(x\in\mathcal{S})\left\{\int_{\mathcal{S}} \frac{\partial\Gamma_\theta(x)}{\partial\theta}\frac{\partial\Gamma_\theta(x)}{\partial\theta'} f(x)\,\mathrm{d}x\right\}^{-1} \frac{\partial\Gamma_\theta(x)}{\partial\theta}\Gamma_f(x). \tag{24}$$

is the leading term that drives the asymptotic distribution of the estimator.

Theorem 8. *Under the null and assumptions A1 to A5, $n^{1/2}\left(\hat{\theta}_f^H - \theta_0\right) \xrightarrow{d} N(0,\Omega_H)$, where $\Omega_H = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left[\vartheta_f^H(x_i), \vartheta_f^H(x_{i+k})\right]$ is the long-run covariance matrix of ϑ_f^H. In case assumption A6 holds, one can employ Newey and West's (1987) non-parametric correction to obtain a consistent estimate of the asymptotic variance.*

4.4 Nuisance parameter result

All results so far consider testing an observable process $\{x_i\}$ with discrete observations $x_1,\ldots,x_n$. In the context of conditional duration models, the interest is in testing the standardised errors $\epsilon_i = x_i/\psi_i$, $i = 1,\ldots,n$. However, the process $\{\epsilon_i\}$ is unobservable and the testing procedure must then proceed using standardised residuals $\hat{\epsilon}_i = x_i/\hat{\psi}_i$, $i = 1,\ldots,n$. In the sequel, we derive conditions in which the H-test is nuisance parameter free, and hence there is no asymptotic cost in substituting standardised residuals for errors. The nuisance parameter result follows in the same line for the D-test, and it is therefore omitted.

To simplify notation, let $e_i = e_i(\phi_0) = x_i/\psi_i(\phi_0)$ and $\hat{e}_i = e_i(\hat{\phi}) = x_i/\psi_i(\hat{\phi})$, where $\hat{\phi}$ is a n^d-consistent estimator of the true parameter ϕ_0. The H-test measures then the closeness between the parametric estimate $\Gamma_{\hat{\theta}}(\hat{e}_i)$ and the non-parametric estimate $\Gamma_{\hat{f}}(\hat{e}_i)$ of the baseline hazard rate function. By definition, a test is nuisance parameter free if the statistic evaluated at $\hat{\phi}$ converges to the same distribution of the statistic evaluated at the true parameter ϕ_0. We must show then that, under the null

$$\Lambda_{\hat{f}}(\hat{\phi}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\hat{e}_i \in \mathcal{S}) \left[\Gamma_{\hat{\theta}}\left(\hat{e}_i\right) - \Gamma_{\hat{f}}\left(\hat{e}_i\right)\right]^2 \tag{25}$$

has the same limiting distribution of its counterpart $\Lambda_{\hat{f}}(\phi_0)$ in (17).

We start by pursuing a third order Taylor expansion with Lagrange remainder of $\Lambda_{\hat{f}}(\hat{\phi})$ about $\Lambda_{\hat{f}}(\phi_0)$, i.e.

$$\begin{aligned}
\Lambda_{\hat{f}}(\hat{\phi}) &= \Lambda_{\hat{f}}(\phi_0) + \Lambda'_{\hat{f}}(\phi_0)(\hat{\phi}-\phi_0) + \frac{1}{2}\Lambda''_{\hat{f}}(\phi_0)(\hat{\phi}-\phi_0,\hat{\phi}-\phi_0) \\
&\quad + \Lambda'''_{\hat{f}}(\phi_*)(\hat{\phi}-\phi_0,\hat{\phi}-\phi_0,\hat{\phi}-\phi_0) \\
&= \Lambda_{\hat{f}}(\phi_0) + \Delta_1 + \Delta_2 + \Delta_3,
\end{aligned}$$

where $\Lambda^{(i)}_{\hat{f}}(\phi_0)$ denotes the i-th order differential of $\Lambda_{\hat{f}}$ with respect to ϕ evaluated at ϕ_0 and $\phi_* \in [\phi_0, \hat{\phi}]$. The first derivative reads

$$\begin{aligned}
\Lambda'_f(\phi_0) &= 2\int_{\mathcal{S}} \left[\Gamma_\theta(e) - \Gamma_f(e)\right]\left[\Gamma'_\theta(e) - \Gamma'_f(e)\right] f(e)\mathrm{d}e \\
&\quad + \int_{\mathcal{S}} \left[\Gamma_\theta(e) - \Gamma_f(e)\right]^2 f'(e)\mathrm{d}e,
\end{aligned} \tag{26}$$

where all differentials are with respect to ϕ evaluated at ϕ_0.

Under the null hypothesis, $\Lambda'_f(\phi_0) = 0$ and $\Lambda'_{\hat{f}}(\phi_0) = O_p\left(n^{-1}h_n^{-1}\right)$ given that $\left(\hat{f}-f\right)^2 = O_p\left(n^{-1}h_n^{-1}\right)$ and $\left(\hat{f}'-f'\right)^2 = O_p\left(n^{-1}h_n^{-3}\right)$. Thus, the first term Δ_1 is of order $O_p\left(n^{-(d+1)}h_n^{-1}\right)$. Similarly, $\Lambda''_{\hat{f}}(\phi_0) = O_p\left(n^{-1}h_n^{-3}\right)$ and $\Delta_2 = O_p\left(n^{-(2d+1)}h_n^{-3}\right)$. The last term requires more caution for it is not evaluated at the true parameter ϕ_0. However, it is not difficult to show that

$$\sup_{|\phi_*-\phi_0|<\epsilon} |\Lambda'''_{\hat{f}}(\phi_*)| = O_p\left(n^{-1/2}h_n^{-7/2}\right) + O_p\left(n^{-1}h_n^{-3}\right), \tag{27}$$

so that $\Delta_3 = O_p\left(n^{-(3d+1/2)}h_n^{-7/2}\right) + O_p\left(n^{-(3d+1)}h_n^{-3}\right)$. The limiting distribution of $\Lambda_{\hat{f}}(\hat{\phi})$ and $\Lambda_{\hat{f}}(\phi_0)$ coincide if and only if

$$nh_n^{1/2}(\Delta_1 + \Delta_2 + \Delta_3) = o_p(1). \tag{28}$$

Under the assumption A4, the bandwidth is of order $o\left(n^{-2/9}\right)$ and hence

$$\begin{aligned}
nh_n^{1/2}\Delta_1 &= o\left(n^{1-1/9}\right)o_p\left(n^{-(d+7/9)}\right) = o_p\left(n^{1/9-d}\right) && (29)\\
nh_n^{1/2}\Delta_2 &= o\left(n^{1-1/9}\right)o_p\left(n^{-(2d+1/3)}\right) = o_p\left(n^{5/9-2d}\right) && (30)\\
nh_n^{1/2}\Delta_3 &= o\left(n^{1-1/9}\right)\left[o_p\left(n^{5/18-3d}\right) + o_p\left(n^{-(3d+1/3)}\right)\right] \\
&= o_p\left(n^{21/18-3d}\right) + o_p\left(n^{5/9-3d}\right), && (31)
\end{aligned}$$

which means that the H-test is nuisance parameter free provided that $d \geq 7/18$. For the gamma kernel version of the H-test, the same argument applies as b_n is of the same order of h_n^2.

5 Numerical results

In this section, we conduct a limited Monte Carlo exercise to assess the performance of our tests in finite samples. The motivation rests on the fact that most non-parametric tests entail substantial size distortions in finite samples. For instance, Fan and Linton (1997) demonstrate how neglecting higher order terms that are close in order to the dominant term may provoke such distortions. Further, despite the results on asymptotic local power, it seems paramount to evaluate the power of our tests against fixed alternatives in finite sample.

The design takes after Grammig and Maurer (1999). We generate 15000 realisations of the linear ACD model of first order, i.e.

$$\psi_i = \omega + \alpha x_{i-1} + \beta\psi_{i-1}, \tag{32}$$

by drawing $e_i = x_i/\psi_i$ from three distributions: exponential, Weibull with $\kappa = 0.6$ and Burr with $\kappa = 2$ and $\sigma^2 = 1.5$. We set $\alpha = 0.1$ and $\beta = 0.7$ to match the typical estimates found in empirical applications. Further, we normalise the unconditional expected duration to one by imposing $\omega = 1-(\alpha+\beta)$ and then set $\psi_0 = 1$ to initialise (32). Along with the full sample ($n = 15000$), we consider a subsample formed by the last 3000 realisations so as to mitigate initial effects. These are typical sample sizes for data on trade and price durations, respectively. All results are based on 1000 replications.

For each replication and data generating process, we first compute maximum likelihood estimates for ACD models with exponential, Weibull and Burr distributions. Optimisation is carried out by taking advantage of Han's (1977) sequential quadratic programming algorithm, which allows for general inequality constraints. Next, we examine the outcomes of our five tests: the D- and H-tests with Gaussian and gamma kernels applied to the standardised residuals and the D-test with Gaussian kernel applied to log-standardised residuals. Bearing in mind assumption A4, we adjust Silverman's (1986) rule of thumb to select the bandwidth h_n for fixed kernel density estimation. The normal distribution serves as reference only for the log-standardised durations, the reference being the exponential otherwise. For simplicity, the gamma kernel density estimation is carried out using $b_n = h_n^2$ as suggested by the asymptotic theory.

The frequency of rejection of the null hypothesis is then computed in order to evaluate size and power of such tests. More precisely, size distortions are investigated by looking at all instances in which the estimated model nests the true specification, e.g. the likelihood considers a Burr density, though the true distribution is exponential or Weibull. Conversely, to investigate the power of these tests, we examine situations in which the estimated model does not encompass the true specification, e.g. the estimated model specify an exponential distribution, whereas the true density is Weibull or Burr.

Figures 1 to 4 display the main results for $n = 3000$ using Davidson and MacKinnon's (1998) graphical representation. Each figure consists of several charts, which are set up in the same way. On the horizontal axe is the significance level and on the vertical axe is the probability of rejection at that significance level. Ideally the size of a test, i.e. the probability of rejection under the null, coincides with the significance level, whereas the power, i.e. the probability of rejection under the alternative, is close to one. To take size distortions into consideration, we consider size-corrected power, i.e. the probability of rejection given simulated rather than asymptotic critical values.

The performance of the D-test for log-standardised durations is a salient feature in all figures. The results are quite encouraging in that such testing procedure is mildly conservative and have excellent power. Besides, the amount of trimming does not seem to affect these results. In fact, no trimming seems the best strategy, though the differences are not statistically significant. On the other hand, the other four tests are to some extent disappointing. In particular, the inferior performance of tests based on gamma kernels are somewhat surprising in view of the absence of boundary bias. Such outcome may be due to the inefficient criterion we have adopted to chose the bandwidth.

Figures 1 and 2 consider the case in which durations follow a Burr ACD process. Figure 1 shows that both D- and H-tests using a Gaussian kernel fail to entail good size performance. In particular, the H-test with Gaussian kernel rejects in every instance the specification of the model, though it is correct. Heavy trimming in the lower tail improves slightly the performance of the D-test, but the distortions are still substantial. Using a gamma kernel, the probability of rejection of the D-test is about 42% irrespective of the weighting scheme and the level of significance at hand. A similar result is due to the H-test with gamma kernel.

Figure 2 illustrates the fact that our tests have, in general, good power against exponential (first column) and Weibull (second column) alternatives. Using a Gaussian kernel, the D-test necessitates heavy trimming in the lower tail, whereas the H-test requires trimming in the upper tail. The intuition is simple. Density estimation with fixed kernels performs poorly close to the origin due to the boundary bias and thus deleting the observations in the lower tail decreases distortions in the D-test. By the same token, pointwise estimates of the hazard rate function are quite unstable in the upper tail because the survival function approaches zero. Therefore, it is not surprising that a higher amount of trimming is necessary in the upper tail for the H-test. Accordingly, the good size-corrected power of both D- and H-tests with no trimming comes at the expense of huge size distortions (see figure 1).

The first and second column of figure 3 document respectively the size and power of our tests when standardised durations have a Weibull distribution. The most striking feature in figure 3 is the complete failure of the D-test with gamma kernel and both H-tests in terms of size performance. In turn, the D-test using a Gaussian kernel performs reasonably well provided that severe trimming is applied to the lower tail; power is trivial otherwise. The intuition is two-fold. First, as aforementioned, this sort of trimming is necessary to counteract the boundary bias of fixed kernel density estimation. Second, the Weibull density is typically very steep near the origin. As durations get close to zero, the parametric estimates of the density approaches infinity as opposed to non-parametric estimates which are bounded. As such, squared differences can get extremely large and the remedy is to introduce more trimming.

Figure 4 reveals that size distortions are less palpable when durations follow an exponential ACD model. The D-test using a Gaussian kernel is slightly more conservative than the D-test applied to log-standardised residuals. Severe trimming in the upper tail is afresh essential to H-tests, though size distortions remain material. Last but not least, our results accord with Grammig and Maurer (1999) in that there is no increase in size distortions if the estimated model considers a more general distribution than necessary. Differences are so minor that we have opted to display only the case in which we estimate a Burr ACD model, though the true distribution is exponential.

To conserve on space, we refrain from displaying similar graphs for the full sample ($n = 15000$) in view that, on balance, the results bear great resemblance. Nonetheless, we collect in table 1 the main statistics for the case in which the data follow a Burr ACD model. In particular, size distortions remain roughly constant, whereas power improves mildly in general – major improvements take place only for the H-tests. In all, the D-test for log-standardised durations seem to outperform the other variants we have proposed. Nonetheless, as the other tests also entail reasonable size-corrected power, one may take advantage of resampling techniques to mitigate size distortions.

6 Empirical application

In this section, we use real world data to test the performance of the linear ACD model (32) with exponential, Weibull and Burr distributions. Data were kindly provided by Luc Bauwens and Pierre Giot and refer to the NYSE's Trade and Quote (TAQ) data set. Bauwens and Giot (1997 and 1998) and Giot (1999) describe more thoroughly the data.

We focus on data ranging from September to November 1996. In particular, we look at price duration processes of five actively traded stocks from the Dow Jones index: Boeing, Coca-Cola, Disney, Exxon, and IBM. Trading at the NYSE is organised 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. Price durations are defined by thinning the quote process with respect to a minimum change in the mid-price of the quotes. We define price duration as the time interval needed to observe a cumulative change in the mid-price of at least \$0.125 as in Giot (1999).

For all stocks, durations between events recorded outside the regular opening hours of the NYSE as well as overnight spells are removed. As documented by Giot (1999), price durations feature a strong time-of-day effect related to predetermined market characteristics such as trade opening and closing times and lunch time for traders. To account for this anomaly, we consider seasonally adjusted price durations $x_i = X_i/\varrho(t_i)$, where X_i is the raw price duration in seconds and $\varrho(\cdot)$ denotes a daily seasonal factor which is determined by averaging durations over thirty minutes intervals for each day of the week and fitting a cubic spline with nodes at each half hour. The resulting (seasonally adjusted) price durations x_i serve then as input in the sequel.

Table 2 reports some descriptive statistics for price durations. There are two common features across stocks: highly significant serial correlation and some degree of overdispersion. That is not surprising: Indeed, ACD models are precisely designed to deal with these stylised facts.

6.1 Estimation and test results

We invoke (quasi) maximum likelihood methods to estimate linear ACD models with exponential, Weibull and Burr distributions. We address both in-sample and out-of-sample performances by splitting the sample. More precisely, we reserve the last third for out-of-sample evaluation. Table 3 summarises the estimation results. For every stock, the Burr ACD model reveals a considerable better fit as indicated by log-likelihoods. On the contrary, the gains in using a Weibull rather than an exponential distribution are quite marginal in most instances. To see why, it suffices to notice that the Weibull estimates of κ are always close to one. In fact, it turns out that $\hat{\kappa} < 1$ for every Weibull ACD model, implying that the hazard rate function decreases monotonically with the standardised duration. Conversely, κ estimates are significantly greater than one for all Burr ACD models, what indicates non-monotonic baseline hazard rate functions. Accordingly, ACD specifications with exponential and Weibull distributions produce similar estimates for duration processes as opposed to Burr ACD models. For Boeing and IBM price durations, differences are indeed striking. All in all, parameter estimates suggest substantial persistence in the rate at which price changes.

Next, we evaluate the performance of the estimated ACD models by examining both in- and out-of-sample standardised durations, which we hereafter refer as residuals and forecast errors, respectively. Tables 4 to 6 portray the results of the D- and H-tests, which are very much in line with Bauwens, Giot, Grammig and Veredas's (2000) analysis rooted in density forecasting techniques. Table 4 reports the p-values of the D-test using a Gaussian kernel for log-standardised durations. As fingered by the Monte Carlo investigation, there is no need for trimming. Residual analysis favours clearly the Burr ACD model as it cannot be rejected at conventional levels of significance for Boeing, Coca-Cola, Disney and Exxon price durations. Contrariwise, the exponential and Weibull alternatives perform quite poorly for every stock, but the Coca-Cola. The linear ACD model is rejected both in- and out-of-sample for IBM price durations irrespective of the distribution. Inspecting the other forecast errors, we find evidence of misspecification only for Boeing and Disney price durations, what probably reflects the presence of structural changes.[1]

Table 5 displays the outcomes of the D-test with Gaussian kernel for raw standardised durations. We consider three weighting strategies. The first exerts no trimming whatsoever, what should produce an extremely conservative test given the results in section 5. Indeed, apart from a borderline result for the Disney residuals of the Burr ACD model, such testing procedure always rejects the null. The second scheme trims realisations out of the interval $(x, 1-x)$, where x denote the empirical 0.025-quantile. As expected, besides some few cases involving residuals of Burr ACD models, rejecting the null remains the rule. Lastly, applying heavy trimming in the lower tail recovers by a long chalk the figures in table 3. The only difference is that the Burr ACD model appear now to produce Boeing forecast errors and IBM residuals that satisfy the null. Of course, this is perchance an artifact due to the weighting procedure since

[1] Further analysis reveal indeed that the last third of the sample yields quite distinct estimates for linear ACD models. Nonetheless, the p-values of the D-test for log-standardised durations depict a pattern similar to previous in-sample results. It easily rejects both exponential and Weibull specifications in every instance, whereas the Burr ACD model fail only for IBM price durations. These additional results are of course available upon request.

misspecification might occur precisely in the trimmed part of the distribution.

Table 6 documents once more how unreliable are H-tests using a Gaussian kernel. Model specification is rejected in nearly all cases even if we introduce severe trimming in the upper tail as suggested in section 5. By the same token, tests based on gamma kernels do not seem very informative. Indeed, all p-values are inferior to 0.0005, mirroring the flimsy finite sample properties of such tests. Figures 5 illustrates the results by plotting the non- and parametric density estimates for Exxon standardised durations. If, on the one hand, non-parametric density estimates oscillate nicely around estimates from the Burr ACD specification; on the other hand, parametric estimates implied by the exponential and Weibull alternatives are consistently above or below their non-parametric counterparts in some intervals.

For completeness, we check whether standardised residuals are serial independent using the BDS test (Brock, Dechert, Scheinkman and LeBaron, 1996). In contrast to the Ljung-Box statistic, the BDS test is sensitive not only to serial correlation but also to other forms of serial dependence. Moreover, the BDS test is nuisance parameter free for additive models (de Lima, 1996), what is quite convenient given that we test estimated residuals rather than true errors. A simple log-transformation renders the linear ACD model additive, hence it suffices to work with log-standardised durations. Table 7 reports the results. For the Boeing price durations, serial independence seems consistent only with the residuals of the Burr ACD model. For Coca-Cola, ACD models seem to produce serially independent residuals irrespective of the distribution, though out-of-sample performances are poor. In turn, all ACD models seem to capture well enough both in- and out-of-sample intertemporal dependence for Disney price durations. Evidence is somewhat inconclusive for Exxon price durations by virtue of the multitude of borderline results. In contrast, the p-values for the IBM log-standardised durations provide strong evidence against the serial independence of both residuals and forecast errors.

Altogether, the figures in table 8 reinforce the evidence provided by the D-test in tables 3 and 4. In particular, none of the linear ACD models seems to fit properly IBM price durations. In turn, the Burr ACD model entails superior performance relative to the exponential and Weibull ACD models for the other four price durations.

7 Concluding remarks

This paper deals with specification tests for conditional duration models, though there is no impediment in using such tests in other contexts. For instance, one could test GARCH-type models by checking whether the distribution of the standardised error is correctly specified. Similarly, Cox's (1955) proportional hazard model implies testable restrictions in the hazard rate function. The main reason to focus on conditional duration models stems from the poor performance of quasi maximum likelihood methods in this context (Grammig and Maurer, 1999).

We propose two testing strategies, namely the D- and H-tests, which rely on gauging the discrepancy between non- and parametric estimates of the density and baseline hazard rate functions of standardised durations, respectively. Asymptotic theory is derived for non-parametric density estimation using both fixed and gamma kernels. The motivation for the latter is to avoid the boundary bias that plagues fixed kernel estimation. All in all, our tests have some attractive theoretical properties. First, they examine the whole distribution of the standardised residuals instead of a limited number of moment restrictions. Second, they are nuisance parameter free. Third, they are suitable to weak dependent time series and, as such, there is no need to test previously for serial independence of the standardised errors.

There are two main topics for future research. First, it is still unclear how to select bandwidths for both fixed and gamma kernel estimations. A possible solution relies on cross-validation methods, which Chen (2000) shows to be particularly valuable to gamma kernel estimation. More precisely, one builds a grid of bandwidth values satisfying assumption A4 and then takes the bandwidth that minimises the test statistic. Second, resampling techniques may deliver more accurate critical values. Indeed, there is vast literature on bootstrapping smoothing-based tests, e.g. Fan (1995) and Li and Wang (1998). Under serial independence of the standardised residuals, the usual bootstrap algorithm presumably works. Suitable bootstrap schemes are also available under weak dependence, such as Politis and Romano's (1994) stationary bootstrap and Bühlmann's (1996) sieve bootstrap, in case one prefers to relax the serial independence assumption.

Appendix: Proofs

Lemma 1. Consider the functional $I_G = \int_0^\infty \varphi_x \tilde{f}_x^2 \, \mathrm{d}x$, where $\tilde{f}_x = \tilde{f}(x)$ is a pointwise gamma kernel estimate of $f_x = f(x)$. Under assumptions A1, A2 and A4,

$$n b_n^{1/4} I_G - \frac{b_n^{-1/4}}{2\sqrt{\pi}} E\left[x^{-1/2}\varphi_x\right] \xrightarrow{d} N\left(0, \frac{1}{\sqrt{\pi}} E\left[x^{-1/2}\varphi_x^2 f_x\right]\right),$$

provided that the above expectations exist.

Proof. See Fernandes (1999).

Lemma 2. Suppose that a functional Φ_f is Fréchet-differentiable relative to the Sobolev norm of order $(2, m)$ at the true density function f with a regular functional derivative ϕ_f. Then, under assumptions A1 to A4, $n^{1/2}\big(\Phi_{\hat{f}} - \Phi_f\big) \xrightarrow{d} N\big(0, V_\Phi\big)$, where $V_\phi = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\big[\phi_f(x_i), \phi_f(x_{i+k})\big]$ is the long run covariance matrix of ϕ_f.

Proof. See Aït-Sahalia (1994).

Lemma 3. Consider a sequence $\{X_i : i = 1, \ldots, n\}$ that satisfies assumption A1. Suppose that the U-statistic $U_n \equiv \sum_{1 \le i < j \le n} H_n(X_i, X_j)$ with symmetric variable function $H_n(\cdot, \cdot)$ is centred and degenerate. If

$$\frac{E_{X_1,X_2}\left\{E_{X_1}^2\left[H_n(X_1,X_1)H_n(X_1,X_2)\right]\right\} + \frac{1}{n} E_{X_1,X_2}\left[H_n^4(X_1,X_2)\right]}{E_{X_1,X_2}^2\left[H_n^2(X_1,X_2)\right]} \longrightarrow 0$$

as sample size grows, then

$$U_n \xrightarrow{d} N\left(0, \frac{n^2}{2} E_{X_1,X_2}\left[H_n^2(X_1,X_2)\right]\right).$$

Proof. See Hall (1984) and Khashimov (1992).

Lemma 4. Consider the functional $I = \int_x \varphi_x \hat{f}_x^2 \, \mathrm{d}x$, where $\int_x$ denotes the integral over the support of x and $\hat{f}_x = \hat{f}(x)$ is a pointwise fixed kernel estimate of $f_x = f(x)$. Under assumptions A1 to A4,

$$n h_n^{1/2} I - h_n^{-1/2} \, e_K \, E\left[\varphi_x\right] \xrightarrow{d} N\left(0, v_K \, E\left[\varphi_x^2 f_x\right]\right),$$

provided that the above expectations are finite.

Proof. The derivation uses lemma 3, i.e. Khashimov's generalisation of Hall's central limit theorem for degenerate U-statistics to weakly dependent stationary processes. We start by decomposing the functional in order to force the emergence of a degenerate U-statistic. Let $r_n(x,X) = \varphi_x^{1/2} K_{h_n}(x-X)$ and $\check{r}_n(x,X) = r_n(x,X) - E_X\big[r_n(x,X)\big]$, where $K_{h_n}(u) = h_n^{-1}K\big(u/h_n\big)$. Then,

$$I = \int_x \left[\sum_{i=1}^n r_n(x,X_i)/T\right]^2 \mathrm{d}x = \frac{1}{n^2}\sum_{i,j}\int_x r_n(x,X_i)r_n(x,X_j)\mathrm{d}x,$$

or equivalently, $I = I_1 + I_2 + I_3 + I_4$, where

$$\begin{aligned}
I_1 &= \frac{2}{n^2}\sum_{i<j}\int_x \check{r}_n(x,X_i)\check{r}_n(x,X_j)\mathrm{d}x \\
I_2 &= \frac{1}{n^2}\sum_{i}\int_x r_n^2(x,X_i)\mathrm{d}x \\
I_3 &= \frac{n(n-1)}{n^2}\int_x E_X^2\left[r_n(x,X)\right]\mathrm{d}x \\
I_4 &= \frac{2n(n-1)}{n^2}\int_x \check{r}_n(x,X_i)E_X\left[r_n(x,X)\right]\mathrm{d}x.
\end{aligned}$$

We show in the sequel that the first term is a degenerate U-statistic and will contribute with the variance in the limiting distribution, whilst the second will contribute with the asymptotic mean. In addition, the third and fourth terms are negligible under assumption A4. The first moment of $r_n(x,X)$ reads

$$\begin{aligned}
E_X\big[r_n(x,X)\big] &= \varphi_x^{1/2}\int_X K_{h_n}(x-X)f(X)\mathrm{d}X = \varphi_x^{1/2}\int_u K(u)f(x+uh_n)\mathrm{d}u \\
&= \varphi_x^{1/2}\int_u K(u)\left[f(x) + \frac{1}{2}f'(x)uh_n + f''(x^*)u^2h_n^2\right]\mathrm{d}u \\
&= \varphi_x^{1/2}f_x + O\left(h_n^2\right),
\end{aligned}$$

where $f^{(i)}$ denotes the i-th derivative of f and $x^* \in [x, x+uh_n]$. Applying similar algebra to the second moment yields $E_X\left[r_n^2(x,X)\right] = h_n^{-1}\, e_K\, \varphi_x f_x + O(1)$. This means that

$$\begin{aligned}
E(I_2) &= \frac{1}{n}\int_x E_X\left[r_n^2(x,X)\right]\mathrm{d}x = \frac{1}{n}\int_x \left[h_n^{-1}\, e_K\, \varphi_x f_x + O(1)\right]\,\mathrm{d}x \\
&= n^{-1}h_n^{-1}\, e_K \int_x \varphi_x f_x\,\mathrm{d}x + O\left(n^{-1}\right),
\end{aligned}$$

whereas $\mathrm{Var}(I_2) = O\left(n^{-3}h_n^{-2}\right)$. It follows then from Chebyshev's inequality that $nh_n^{1/2}I_2 - h_n^{-1/2}\, e_K\, E[\varphi_x] = o_p(1)$. In turn, we have that

$$I_4 = \frac{n(n-1)}{n^2}\int_x E_X^2\left[r_n(x,X)\right]\,\mathrm{d}x = \frac{n(n-1)}{n^2}O\left(h_n^4\right) = O\left(h_n^4\right),$$

which, under assumption A4, implies that $nh_n^{1/2}I_4 = o(1)$. Further,

$$E(I_3) = \frac{2(n-1)}{n}\int_x E_X\left[\check{r}_n(x,X)\right]\, E_X\big[r_n(x,X)\big]\,\mathrm{d}x = 0,$$

whilst $E(I_3^2) = O\left(n^{-1}h_n^4\right)$. It suffices then to impose assumption A4 to ensure, by Chebyshev's inequality, that $nh_n^{1/2}I_3 = o_p(1)$. Finally, recall that $I_1 = \sum_{i<j} H_n(X_i, X_j)$, where $H_n(X_i, X_j) = 2n^{-2}\int_x \check{r}_n(x, X_i)\check{r}_n(x, X_j)\mathrm{d}x$. As $H_n(X_i, X_j)$ is symmetric, centred and such that $E\left[H_n(X_i, X_j)\middle|X_j\right] = 0$ almost surely, I_1 is a degenerate U-statistic. Thus, it follows immediately from lemma 3 that $nh_n^{1/2}I_1 \xrightarrow{d} N(0, V_H)$, where

$$
\begin{aligned}
V_H &= \frac{n^4 h_n}{2} E_{X_1,X_2}\left[H_n^2(X_1, X_2)\right] \\
&= 2h_n \int_{X_1,X_2} \left[\int_x \check{r}_n(x, X_1)\check{r}_n(x, X_2)\mathrm{d}x\right]^2 f(X_1, X_2)\mathrm{d}(X_1, X_2) \\
&= 2h_n \int_{x,y} \left[\int_X \check{r}_n(x, X)\check{r}_n(y, X)f(X)\mathrm{d}X\right]^2 \mathrm{d}(x, y) \\
&= 2\int_{x,v} \varphi_x^2 \left[\int_u K(u)K(u+v)f(x-uh_n)\,\mathrm{d}u\right. \\
&\qquad \left. - h_n \int_u K(u)f(x-uh_n)\,\mathrm{d}u \int_u K(u)f(x+vh_n-uh_n)\,\mathrm{d}u\right]^2 \mathrm{d}(x, v) \\
&\sim 2\int_{x,v} \varphi_x^2 \left[\int_u K(u)K(u+v)f(x-uh_n)\right]^2 \mathrm{d}(x, v) \\
&\sim 2\,v_K \int_x \varphi_x^2\, f_x^2\,\mathrm{d}x,
\end{aligned}
$$

which completes the proof.

Proof of (10). Consider the following expansion

$$
\Psi_{f,h}(\gamma) = \Psi_{f+\gamma h} = \int_S \left[f(x,\theta_\gamma) - f(x) - \gamma h(x)\right]^2 \left[f(x) + \gamma h(x)\right]\mathrm{d}x,
$$

where $\theta_\gamma = \theta_{f+\gamma h}$. Differentiating with respect to γ yields

$$
\begin{aligned}
\frac{\partial \Psi_{f,h}(\gamma)}{\partial\gamma} &= 2\int_S \frac{\partial f(x,\theta_\gamma)}{\partial\theta}\frac{\partial\theta_\gamma}{\partial\gamma}\left[f(x,\theta_\gamma) - f(x) - \gamma h(x)\right]\left[f(x) + \gamma h(x)\right]\mathrm{d}x \\
&\quad - 2\int_S \left[f(x,\theta_\gamma) - f(x) - \gamma h(x)\right]\left[f(x) + \gamma h(x)\right]h(x)\,\mathrm{d}x \\
&\quad + \int_S \left[f(x,\theta_\gamma) - f(x) - \gamma h(x)\right]^2 h(x)\,\mathrm{d}x.
\end{aligned}
$$

Under the null, the parametric specification of the density function is correctly specified, i.e. $f(x,\theta) = f(x)$; hence the first functional derivative $\mathrm{D}\Psi_f = \frac{\partial}{\partial\gamma}\Psi_{f,h}(0)$ is singular. In turn, the second functional derivative reads

$$
\begin{aligned}
\frac{\partial^2 \Psi_{f,h}(\gamma)}{\partial\gamma\partial\gamma'} &= 2\int_S \frac{\partial^2 f(x,\theta_\gamma)}{\partial\theta\partial\theta'}\frac{\partial\theta_\gamma}{\partial\gamma}\frac{\partial\theta_\gamma}{\partial\gamma'}\left[f(x,\theta_\gamma) - f_x - \gamma h_x\right]\left[f_x + \gamma h_x\right]\mathrm{d}x \\
&\quad + 2\int_S \frac{\partial f(x,\theta_\gamma)}{\partial\theta}\frac{\partial^2\theta_\gamma}{\partial\gamma\partial\gamma'}\left[f(x,\theta_\gamma) - f_x - \gamma h_x\right]\left[f_x + \gamma h_x\right]\mathrm{d}x
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\int_S \frac{\partial f(x,\theta_\gamma)}{\partial \theta}\frac{\partial f(x,\theta_\gamma)}{\partial \theta'}\frac{\partial \theta_\gamma}{\partial \gamma}\frac{\partial \theta_\gamma}{\partial \gamma'}[f_x+\gamma h_x]\,\mathrm{d}x \\
&- 4\int_S \frac{\partial f(x,\theta_\gamma)}{\partial \theta}\frac{\partial \theta_\gamma}{\partial \gamma}[f_x+\gamma h_x]\,h_x\,\mathrm{d}x \\
&+ 4\int_S \frac{\partial f(x,\theta_\gamma)}{\partial \theta}\frac{\partial \theta_\gamma}{\partial \gamma}[f(x,\theta_\gamma)-f_x-\gamma h_x]\,h_x\,\mathrm{d}x, \\
&+ 2\int_S [f_x+\gamma h_x]\,h_x^2\,\mathrm{d}x - 4\int_S [f(x,\theta_\gamma)-f_x-\gamma h_x]\,h_x^2\,\mathrm{d}x,
\end{aligned}
$$

which reduces to (10) by evaluating at $\gamma = 0$ and imposing the null.

Proof of Theorem 2. Under the null, the following functional Taylor expansion is valid

$$
\Psi_{f+h} = \int_{x,y} \mathbb{1}(x\in\mathcal{S})\left[\ell_f^D(x,y) + f_x\delta_{(x)}(y)\right]\mathrm{d}H(x)H(y) + O\left(\|h_x\|^3\right),
$$

where ℓ_f^D is a continuous functional which includes the first and second terms of (10) as well as the regular part of its third term and $\delta_{(x)}$ is a Dirac mass at x. Replacing h_x by $\tilde{f}_x - f_x$ ensues that the first term

$$
\int_{x,y} \mathbb{1}(x\in\mathcal{S})\ell_f^D(x,y)\,\mathrm{d}H(x)\mathrm{d}H(y)
$$

is negligible since it converges at a faster rate T to a sum of independent χ^2 distributions (Serfling, 1980; Aït-Sahalia, 1994). In turn, applying lemma 1 with $\varphi_x = \mathbb{1}(x\in\mathcal{S})f_x$ yields that

$$
\int_{x,y} \mathbb{1}(x\in\mathcal{S})f_x\delta_{(x)}(y)\,\mathrm{d}H(x)\mathrm{d}H(y) = \int_S f_x h_x^2\,\mathrm{d}x
$$

converges in distribution at rate $nb_n^{1/4}$ to a Gaussian variate with mean $b_n^{-1/4}\delta_G$ and variance σ_G^2.

Proof of Theorem 3. The conditions imposed are such that the functional Taylor expansion under consideration is valid even in case the x_{in}, $i = 1,\ldots,n$, are a double array. Thus, for the D-test with fixed kernel, it ensues that, under H_{1n}^D and assumptions A1 to A4,

$$
\hat{\tau}_n^D - \frac{nh_n^{1/2}}{\hat{\sigma}_D}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(x_{in}\in\mathcal{S})\left[f(x_{in},\theta_f) - f(x_{in})\right]^2 \xrightarrow{d^{[n]}} N(0,1),
$$

where the superscript $[n]$ denotes dependence on $f^{[n]}$. The first result follows then by noting that $\hat{\sigma}_D \xrightarrow{p^{[n]}} \sigma_D$ and

$$
\Psi_{f^{[n]}} \quad = \quad \frac{1}{n}\sum_{i=1}^n \mathbb{1}(x_{in}\in\mathcal{S}\left[f^{[n]}(x_{in},\theta_{f^{[n]}}) - f^{[n]}(x_{in})\right]^2
$$

$$
\begin{aligned}
&= E\left\{\mathbb{1}(x_{1n}\in\mathcal{S})\left[f^{[n]}(x_{1n},\theta_{f^{[n]}})-f^{[n]}(x_{1n})\right]^2\right\}+O_p\left(n^{-1/2}\right)\\
&= \varepsilon_n^2 E\left[\mathbb{1}(x_{1n}\in\mathcal{S})\ell_D^2(x_{1n})\right]+o_p\left(n^{-1}h_n^{-1/2}\right)\\
&= n^{-1}h_n^{-1/2}\ell_D^{\mathcal{S}}+o_p\left(n^{-1}h_n^{-1/2}\right).
\end{aligned}
$$

Applying a similar argument to the gamma kernel version of the D-test completes the proof (see the proof of theorem 7).

Proof of (18). Consider the following expansion

$$
\Lambda_{f,h}(\gamma)=\Lambda_{f+\gamma h}=\int_{\mathcal{S}}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]^2\left[f(x)+\gamma h(x)\right]\mathrm{d}x,
$$

where $\theta_\gamma=\theta_{f+\gamma h}$ to simplify notation. Differentiating with respect to γ entails

$$
\begin{aligned}
\frac{\partial\Lambda_{f,h}(\gamma)}{\partial\gamma} &= 2\int_{\mathcal{S}}\frac{\partial\Gamma_{\theta_\gamma}(x)}{\partial\theta}\frac{\partial\theta_\gamma}{\partial\gamma}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad-2\int_{\mathcal{S}}\frac{\partial\Gamma_{f+\gamma h}(x)}{\partial\gamma}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad+\int_{\mathcal{S}}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]^2 h(x)\,\mathrm{d}x,
\end{aligned}
$$

which recovers (18) if evaluated at $\gamma=0$.

Proof of (20). Computing the second differential of the expression above with respect to γ yields

$$
\begin{aligned}
\frac{\partial^2\Lambda_{f,h}(\gamma)}{\partial\gamma\partial\gamma'} &= 2\int_{\mathcal{S}}\frac{\partial^2\Gamma_{\theta_\gamma}(x)}{\partial\theta\partial\theta'}\frac{\partial\theta_\gamma}{\partial\gamma}\frac{\partial\theta_\gamma}{\partial\gamma'}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad+2\int_{\mathcal{S}}\frac{\partial\Gamma_{\theta_\gamma}(x)}{\partial\theta}\frac{\partial^2\theta_\gamma}{\partial\gamma\partial\gamma'}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad+2\int_{\mathcal{S}}\frac{\partial\Gamma_{\theta_\gamma}(x)}{\partial\theta}\frac{\partial\Gamma_{\theta_\gamma}(x)}{\partial\theta'}\frac{\partial\theta_\gamma}{\partial\gamma}\frac{\partial\theta_\gamma}{\partial\gamma'}\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad-4\int_{\mathcal{S}}\frac{\partial\Gamma_{\theta_\gamma}(x)}{\partial\theta}\frac{\partial\theta_\gamma}{\partial\gamma}\frac{\partial\Gamma_{f+\gamma h}(x)}{\partial\gamma}\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad+4\int_{\mathcal{S}}\frac{\partial\Gamma_{\theta_\gamma}(x)}{\partial\theta}\frac{\partial\theta_\gamma}{\partial\gamma}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]h(x)\,\mathrm{d}x\\
&\quad-2\int_{\mathcal{S}}\frac{\partial^2\Gamma_{f+\gamma h}(x)}{\partial\gamma\gamma'}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad+2\int_{\mathcal{S}}\frac{\partial\Gamma_{f+\gamma h}}{\partial\gamma}\frac{\partial\Gamma_{f+\gamma h}}{\partial\gamma'}\left[f(x)+\gamma h(x)\right]\mathrm{d}x\\
&\quad-4\int_{\mathcal{S}}\frac{\partial\Gamma_{f+\gamma h}(x)}{\partial\gamma}\left[\Gamma_{\theta_\gamma}(x)-\Gamma_{f+\gamma h}(x)\right]h(x)\,\mathrm{d}x,
\end{aligned}
$$

which equals (20) for $\gamma=0$.

Proof of Theorem 5. Under the null, the following functional Taylor expansion is valid

$$\Lambda_{f+h} = \int_{x,y} \mathbb{1}(x \in \mathcal{S}) \left[\ell_f^H(x,y) + S_x^{-1}\delta_{(x)}(y)\right] \mathrm{d}H(x)H(y) + O\left(\|h\|^3\right),$$

where ℓ_f^H is a continuous functional encompassing the second and third terms of (20) as well as the regular part of its first term and S_x denotes the survival function $1 - F(x)$. Replacing h_x by $\hat{f}_x - f_x$ ensues that the first term

$$\int_{x,y} \mathbb{1}(x \in \mathcal{S})\ell_f^H(x,y)\,\mathrm{d}H(x)\mathrm{d}H(y)$$

converges at a rate T and therefore it is negligible. In turn, applying lemma 4 with $\varphi_x = \mathbb{1}(x \in \mathcal{S})S_x^{-1}$ yields that

$$\int_{x,y} \mathbb{1}(x \in \mathcal{S})S_x^{-1}\delta_{(x)}(y)\,\mathrm{d}H(x)\mathrm{d}H(y) = \int_{\mathcal{S}} S_x^{-1}h_x^2\,\mathrm{d}x$$

converges weakly at rate $nh_n^{1/2}$ to a normal distribution with mean $h_n^{-1/2}\lambda_H$ and variance ς_H^2.

Proof of Theorem 6. Consider the above functional Taylor expansion with $h_x = \tilde{f}_x - f_x$. Once more, the first term converges at a rate T, whereas lemma 1 implies that

$$\int_{x,y} \mathbb{1}(x \in \mathcal{S})S_x^{-1}\delta_{(x)}(y)\,\mathrm{d}H(x)\mathrm{d}H(y) = \int_x \mathbb{1}(x \in \mathcal{S})f_x h_x^2\,\mathrm{d}x$$

converges in distribution at rate $nb_n^{1/4}$ to a normal variate with mean $b_n^{-1/4}\lambda_G$ and variance ς_G^2.

Proof of Theorem 7. Afresh, the corresponding functional Taylor expansion is consistent with the double array sequence x_{in}, $i = 1, \ldots, n$. Thus, for the H-test with gamma kernel, we have that, under H_{1n}^H and assumptions A1 to A4,

$$\tilde{\tau}_n^H - \frac{nb_n^{1/4}}{\hat{\varsigma}_G}\frac{1}{n}\sum_{i=1}^n \mathbb{1}(x_{in} \in \mathcal{S})\left[\Gamma(x,\theta_f) - \Gamma_f(x)\right]^2 \xrightarrow{d^{[n]}} N(0,1).$$

The result follows then from the fact that $\tilde{\varsigma}_G \xrightarrow{p^{[n]}} \varsigma_G$ and

$$\begin{aligned}
\Lambda_{f^{[n]}} &= \frac{1}{n}\sum_{i=1}^n \mathbb{1}(x_{in} \in \mathcal{S}\left[\Gamma^{[n]}(x_{in},\theta_{f^{[n]}}) - \Gamma_f^{[n]}(x_{in})\right]^2 \\
&= E\left\{\mathbb{1}(x_{1n} \in \mathcal{S})\left[\Gamma^{[n]}(x_{1n},\theta_{f^{[n]}}) - \Gamma_f^{[n]}(x_{1n})\right]^2\right\} + O_p\left(n^{-1/2}\right) \\
&= \varepsilon_n^2 E\left[\mathbb{1}(x_{1n} \in \mathcal{S})\ell_H^2(x_{1n})\right] + o_p\left(n^{-1}b_n^{-1/4}\right) \\
&= n^{-1}b_n^{-1/4}\ell_H^{\mathcal{S}} + o_p\left(n^{-1}b_n^{-1/4}\right).
\end{aligned}$$

We omit the proof for the fixed kernel version of the H-test in view that it is completely analogous (see the proof of theorem 3).

Proof of Theorem 8. The implicit functional corresponding the M-estimator associated with the H-test is

$$\int_S \frac{\partial \Gamma(x,\theta_f^H)}{\partial \theta}\left[\Gamma(x,\theta_f^H)-\Gamma_f(x)\right] f(x)\,\mathrm{d}x \equiv 0,$$

which results in the following expansion

$$\int_S \frac{\partial \Gamma(x,\theta_\gamma^H)}{\partial \theta}\left[\Gamma(x,\theta_\gamma^H)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x \equiv 0.$$

Differentiating with respect to γ entails then

$$\begin{aligned}
\int_S \frac{\partial^2 \Gamma(x,\theta_\gamma^H)}{\partial\theta\partial\theta'}\frac{\partial\theta_\gamma^H}{\partial\gamma}\left[\Gamma(x,\theta_\gamma^H)-\Gamma_{f+\gamma h}(x)\right]\left[f(x)+\gamma h(x)\right]\mathrm{d}x \\
+\int_S \frac{\partial \Gamma(x,\theta_\gamma^H)}{\partial\theta}\frac{\partial \Gamma(x,\theta_\gamma^H)}{\partial\theta'}\frac{\partial\theta_\gamma^H}{\partial\gamma}\left[f(x)+\gamma h(x)\right]\mathrm{d}x \\
+\int_S \frac{\partial \Gamma(x,\theta_\gamma^H)}{\partial\theta}\left[\Gamma(x,\theta_\gamma^H)-\Gamma_{f+\gamma h}(x)\right] h(x)\,\mathrm{d}x \\
-\int_S \frac{\partial \Gamma(x,\theta_\gamma^H)}{\partial\theta}\frac{\partial \Gamma_{f+\gamma h}(x)}{\partial\gamma}\left[f(x)+\gamma h(x)\right]\mathrm{d}x &= 0,
\end{aligned}$$

which recovers (23) if one imposes the correct specification of the model and evaluates at $\gamma=0$. As the first term in the right-hand side of (19) converges at a slower rate than the second, (24) will drive the asymptotic distribution of $\theta_{\hat{f}}^H$. A straightforward application of lemma 2 completes then the proof.

References

Aït-Sahalia, Y. (1994). The delta method for nonparametric kernel functionals, Graduate School of Business, University of Chicago.

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate, *Review of Financial Studies* **9**: 385–426.

Bauwens, L. and Giot, P. (1997). The logarithmic ACD model: An application to market microstructure and NASDAQ, *Annales d'Economie et de Statistique*. Forthcoming.

Bauwens, L. and Giot, P. (1998). Asymmetric ACD models: Introducing price information in ACD models with a two-state transition model, Université Catholique de Louvain.

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

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

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Annals of Statistics* **1**: 1071–1095.

Brock, W. A., Dechert, W. D., Scheinkman, J. A. and LeBaron, B. (1996). A test for independence based on the correlation dimension, *Econometric Reviews* **15**: 197–235.

Bühlmann, P. (1996). Sieve bootstrap for time series, Department of Statistics, University of California at Berkeley.

Carrasco, M. and Chen, X. (1999). β-mixing and moment properties of various GARCH, stochastic volatility and ACD models, *CREST Working Paper 9941*, INSEE.

Chen, S. X. (2000). Gamma kernel estimators for density functions, *Annals of the Institute of Statistical Mathematics*. Forthcoming.

Chesher, A., Dhaene, G., Gouriéroux, C. and Scaillet, O. (1999). Bartlett identities tests, *CORE Discussion Paper 9939*, Université Catholique de Louvain.

Cox, D. R. (1955). Some statistical methods connected with series of events (with discussion), *Journal of the Royal Statistical Society B* **17**: 129–164.

Davidson, R. and MacKinnon, J. (1998). Graphical methods for investigating the size and power of hypothesis test, *Manchester School of Economics and Social Studies* **66**: 1–26.

de Lima, P. (1996). Nuisance parameter free properties of correlation integral based statistics, *Econometric Reviews* **15**: 237–259.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**: 863–883.

Engle, R. F. (1996). The econometrics of ultra-high frequency data, *Econometrica*. Forthcoming.

Engle, R. F. and Lange, J. (1997). Measuring, forecasting and explaining time varying liquidity in the stock market, *Economics Discussion Paper 9712*, University of Califonia at San Diego.

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

Engle, R. F. and 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**: 187–212.

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

Fan, Y. (1995). Bootstrapping a consistent nonparametric goodness-of-fit test, *Econometric Reviews* **14**: 367–382.

Fan, Y. and Linton, O. (1997). Some higher order theory for a consistent nonparametric model specification, University of Windsor and Yale University.

Fernandes, M. (1999). Central limit theorem for asymmetric kernel functionals, European University Institute.

Ghysels, E. and Jasiak, J. (1998a). GARCH for irregularly spaced financial data: The ACD-GARCH model, *Studies in Nonlinear Economics and Econometrics* **2**: 133–149.

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

Ghysels, E., Gouriéroux, C. and Jasiak, J. (1997). Stochastic volatility duration models, *CREST Working Paper 9746*, INSEE.

Giot, P. (1999). Time transformations, intraday data and volatility models, *CORE Discussion Paper 9944*, Université Catholique de Louvain.

Gouriéroux, C., Jasiak, J. and Le Fol, G. (1996). Intra-day market activity, *CREST Working Paper 9636*, INSEE.

Grammig, J. and Maurer, K.-O. (1999). Non-monotonic hazard functions and the autoregressive conditional duration model, *SFB373 Discussion Paper 55*, Humboldt-Universität zu Berlin.

Grammig, J. and Wellner, M. (1999). Modeling the interdependence of volatility and inter-transaction duration processes, *SFB373 Discussion Paper 21*, Humboldt-Universität zu Berlin.

Grammig, J., Hujer, R., Kokot, S. and 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, *SFB373 Discussion Paper 55*, Humboldt-Universität zu Berlin.

Hall, P. (1984). Central limit theorem for integrated squared error multivariate nonparametric density estimators, *Journal of Multivariate Analysis* **14**: 1–16.

Hamilton, J. D. and Jorda, O. (1999). A model for the federal funds rate target, University of California at San Diego and University of Califonia at Davis.

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

Hausman, J. (1978). Specification tests in econometrics, *Econometrica* **46**: 1251–1272.

Khashimov, S. A. (1992). Limiting behaviour of generalized U-statistics of weakly dependent stationary processes, *Theory of Probability and Its Applications* **37**: 148–150.

Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for parametric regression function, *Journal of Econometrics* **87**: 147–165.

Lo, A. W., MacKinlay, A. C. and Zhang, J. (1997). Econometric models of limit-order executions, *LFE Working Paper 1028*, Sloan School of Management, MIT.

Lunde, A. (1999a). A conjugate gamma model for durations in transaction data, University of Aarhus.

Lunde, A. (1999b). A generalized gamma autoregressive conditional duration model, University of Aarhus.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator, *Econometric Theory* **10**: 233–253.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**: 703–708.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap, *Journal of the American Statistical Association* **89**: 1303–1313.

Pringent, J.-L., Renault, E. and Scaillet, O. (1999). An autoregressive conditional binomial option pricing model, Université de Cergy-Pontoise, CREST, and Université Catholique de Louvain.

Russell, J. R. (1998). Econometric modeling of multivariate irregularly-spaced high-frequency data, Graduate School of Business, University of Chicago.

Russell, J. R. and Engle, R. F. (1998). Econometric analysis of discrete-valued irregularly-spaced financial transaction data using a new autoregressive conditional multinomial model, *Discussion Paper 9810*, University of Califonia at San Diego.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Wolak, F. (1991). The local nature of hypothesis tests involving inquality constraints in nonlinear models, *Econometrica* **59**: 981–995.

Zhang, M. Y., Russell, J. R. and Tsay, R. S. (1999). A nonlinear autoregressive conditional duration model with applications to financial transaction data, Graduate School of Business, University of Chicago.

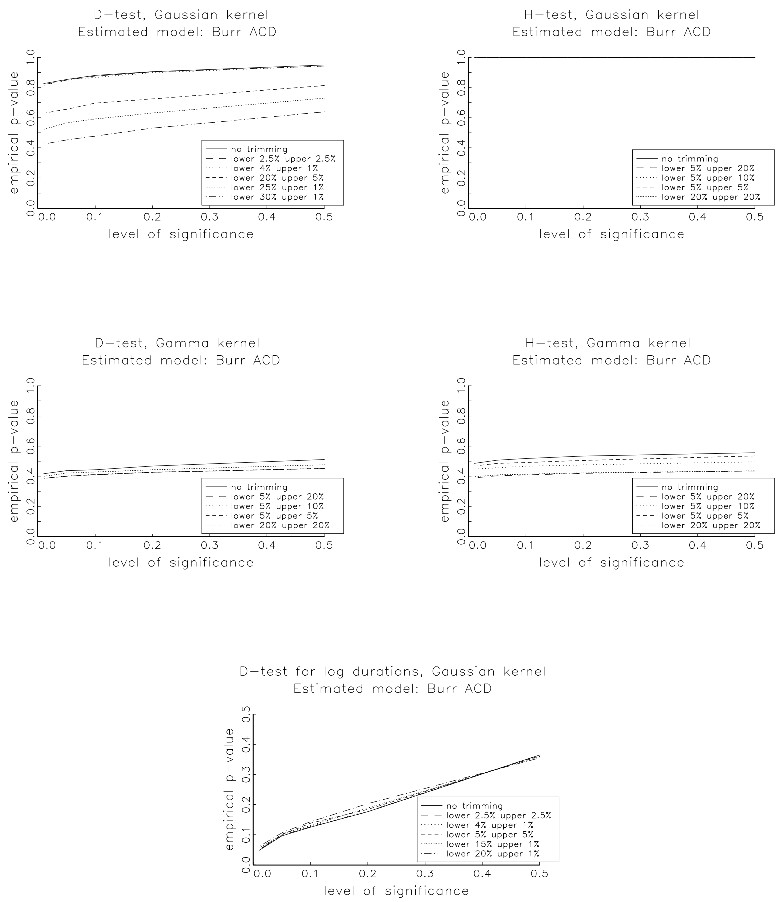


Figure 1: Empirical size, Burr ACD

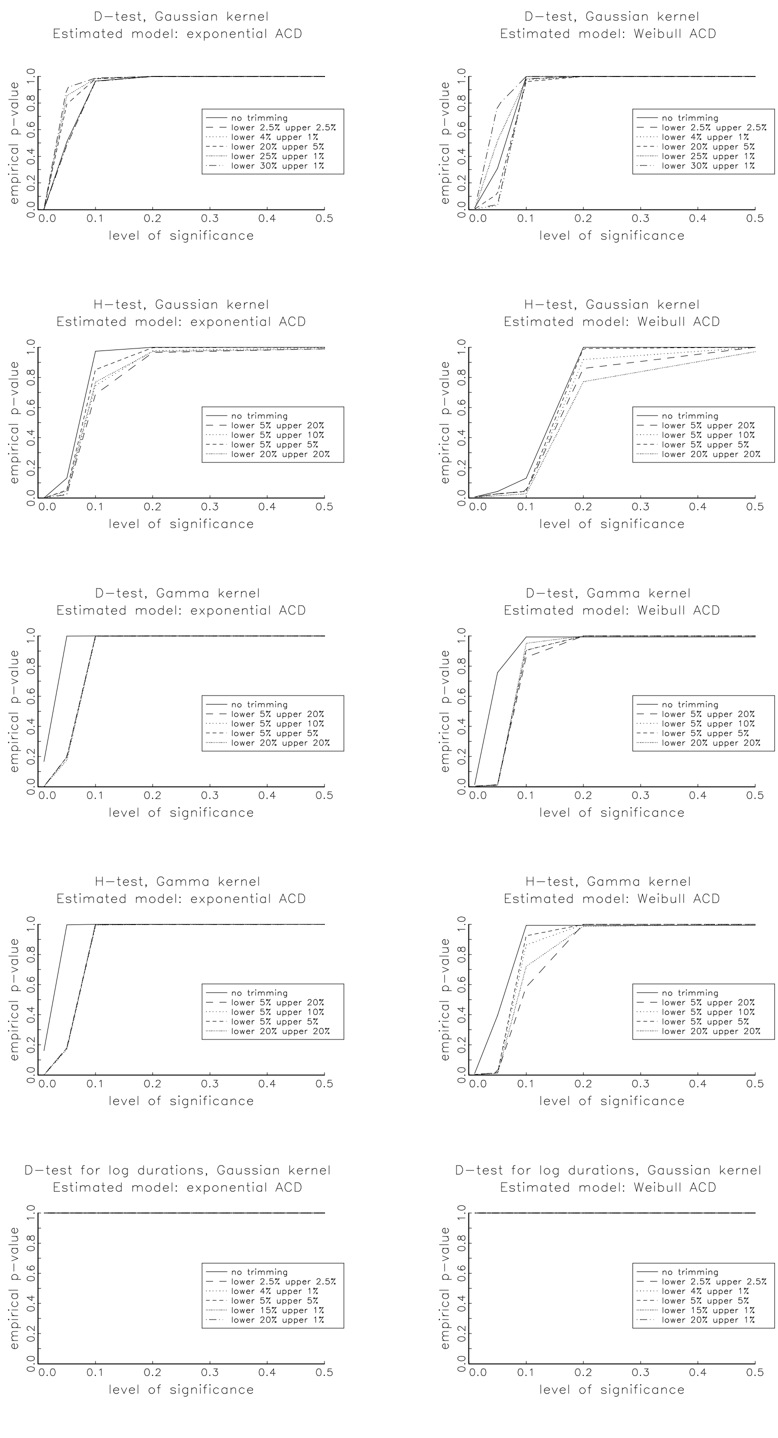


Figure 2: Power against exponential and Weibull alternatives, Burr ACD

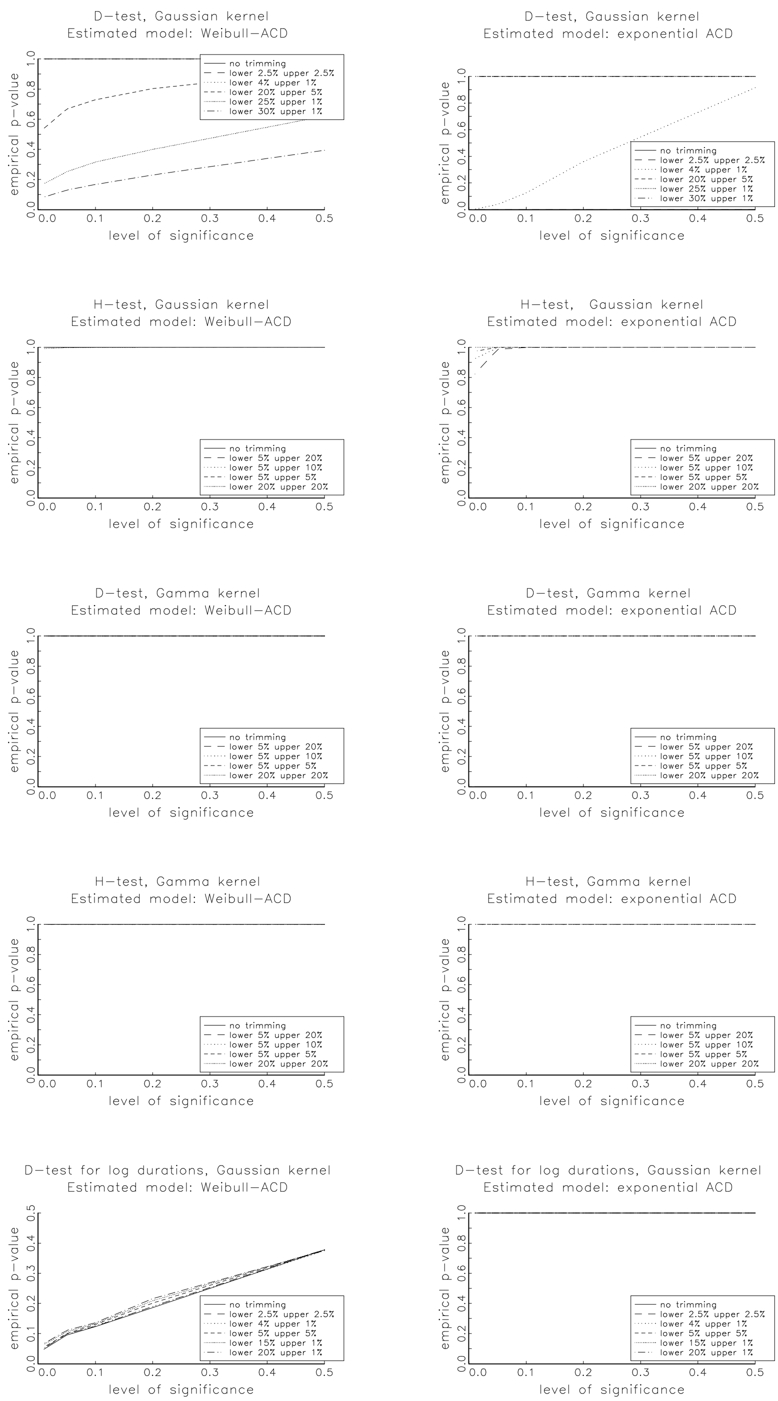


Figure 3: Size and power against exponential alternative, Weibull ACD

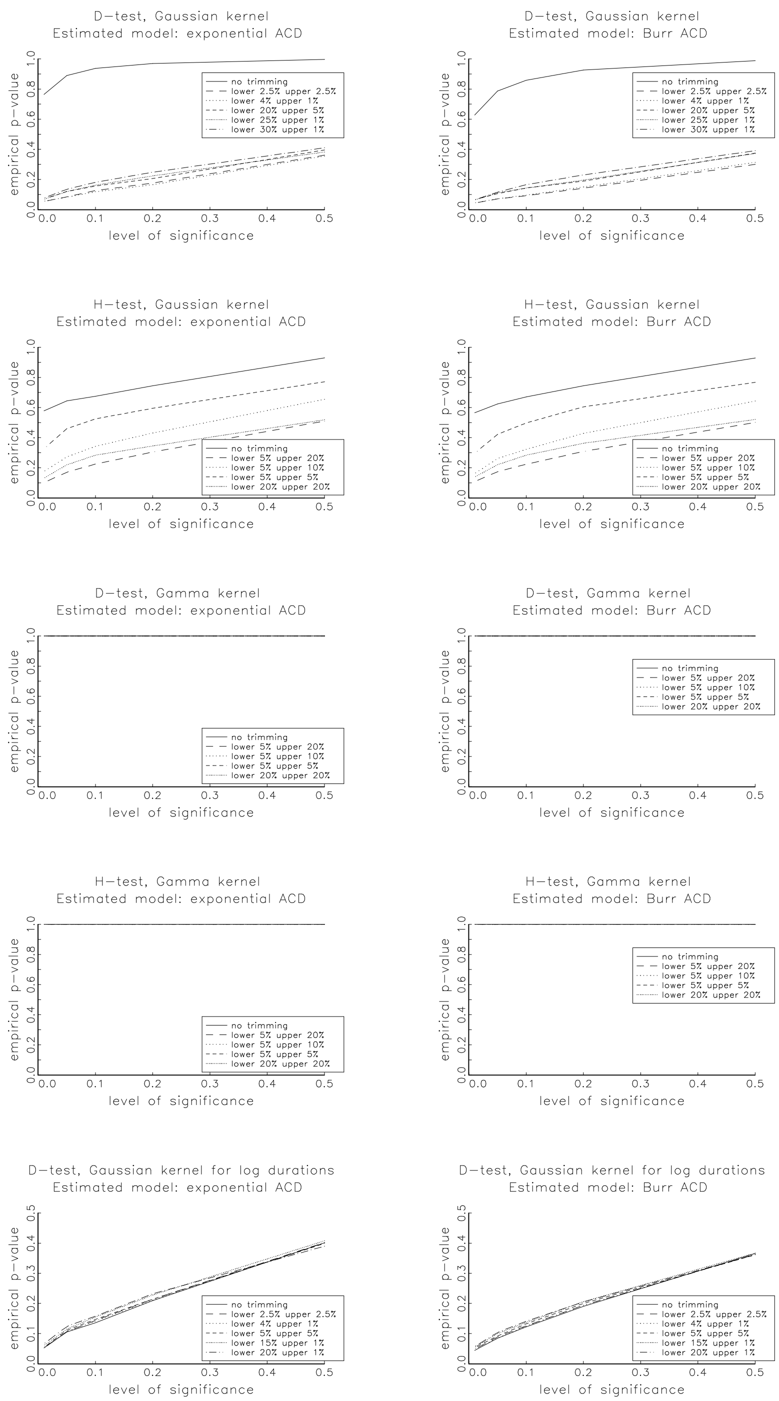
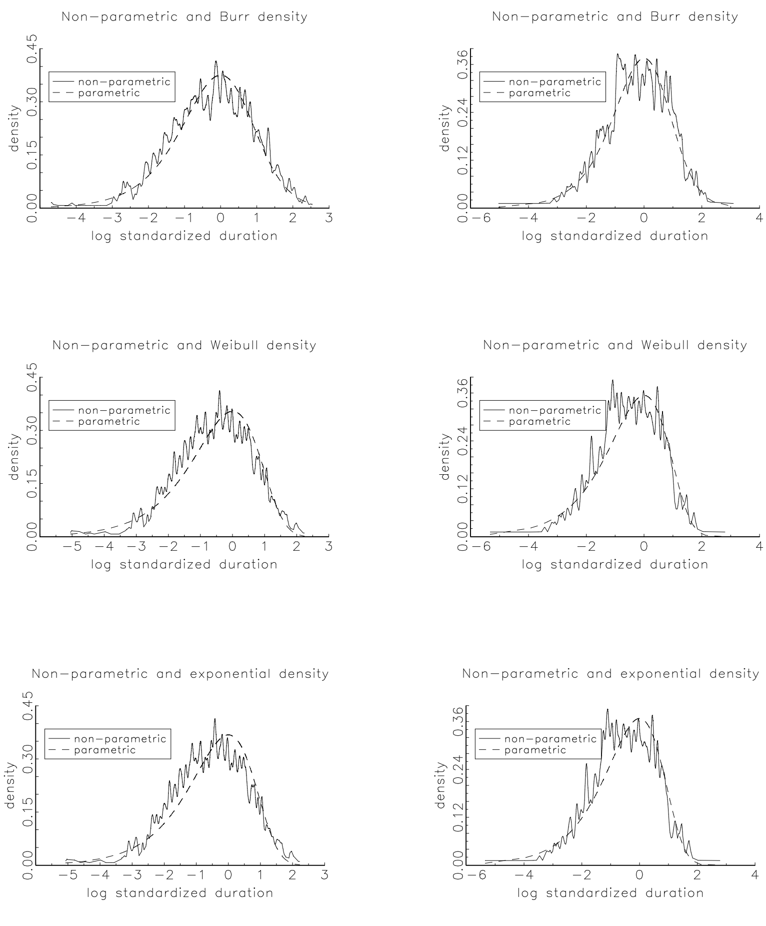


Figure 4: Empirical size, Exponential ACD



(a) in-sample (b) out-of-sample

Figure 5: Exxon price durations

Table 1

Finite sample properties of the testing procedures

Data generating mechanism: Burr ACD process

Sample size: 15000

Number of replications: 1000

trimming	actual size			power: Weibull ACD		
	1%	5%	10%	1%	5%	10%
D-test for standardised durations, Gaussian kernel						
none	0.9720	0.9800	0.9870	0.0130	0.9330	1.0000
(0.025, 0.025)	0.9650	0.9750	0.9840	0.0070	0.1730	1.0000
(0.04, 0.01)	0.9630	0.9720	0.9820	0.0060	0.0530	1.0000
(0.25, 0.01)	0.6980	0.7290	0.7470	0.0020	0.9400	1.0000
(0.30, 0.01)	0.5720	0.6110	0.6500	0.0020	0.9670	1.0000
D-test for log-standardised durations, Gaussian kernel						
none	0.0390	0.0880	0.1120	1.0000	1.0000	1.0000
(0.025, 0.025)	0.0450	0.0900	0.1260	1.0000	1.0000	1.0000
(0.05, 0.05)	0.0470	0.1010	0.1280	1.0000	1.0000	1.0000
D-test for standardised durations, gamma kernel						
none	0.6700	0.6960	0.7110	0.1520	0.9910	0.9910
(0.05, 0.05)	0.6150	0.6380	0.6550	0.0020	0.0110	1.0000
(0.05, 0.20)	0.5820	0.5980	0.6080	0.0020	0.0110	1.0000
(0.20, 0.20)	0.5970	0.6130	0.6260	0.0020	0.0430	0.9850
H-test for standardised durations, Gaussian kernel						
none	0.9570	0.9860	0.9950	0.0090	0.0450	0.9700
(0.025, 0.025)	0.8710	0.9000	0.9110	0.0070	0.0300	0.2460
(0.04, 0.01)	0.8960	0.9310	0.9490	0.0070	0.0270	0.1440
(0.25, 0.01)	0.8930	0.9240	0.9380	0.0030	0.0150	0.6160
(0.30, 0.01)	0.8760	0.8990	0.9150	0.0030	0.0140	0.7980
H-test for standardised durations, gamma kernel						
none	0.9570	0.9860	0.9950	0.0080	0.9910	0.9910
(0.05, 0.05)	0.8000	0.8170	0.8260	0.0030	0.0140	0.9950
(0.05, 0.20)	0.5630	0.5720	0.5780	0.0030	0.0130	0.9870
(0.20, 0.20)	0.5680	0.5740	0.5790	0.0020	0.0100	0.9780

Table 2
Descriptive statistics of price durations

stock	sample size	mean	overdispersion	$Q(10)$
Boeing	2620	1.001	1.338	322.3
Coca-Cola	1609	1.002	1.171	69.7
Disney	2160	1.004	1.209	137.3
Exxon	2717	1.000	1.196	68.2
IBM	6728	1.015	1.427	1932.6

Data correspond to seasonally adjusted durations between bid-ask quotes such that a cumulative change in the mid-price of at least \$0.125 is observed. Overdispersion stands for the ratio between standard deviation and mean. $Q(10)$ denotes the Ljung-Box statistic of order 10.

Table 3
Maximum likelihood estimates of the ACD models

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

The column $\log \mathcal{L}$ displays the value of the log-likelihood, whereas the rows with figures in parentheses report robust standard errors.

Table 4

D-test results for price log-durations, Gaussian kernel

stock	in sample	out of sample
Boeing		
Exponential	0.000	0.000
Weibull	0.000	0.000
Burr	0.138	0.009
Coca-Cola		
Exponential	0.029	0.821
Weibull	0.316	0.877
Burr	0.666	0.969
Disney		
Exponential	0.000	0.000
Weibull	0.000	0.000
Burr	0.160	0.000
Exxon		
Exponential	0.000	0.007
Weibull	0.000	0.028
Burr	0.137	0.261
IBM		
Exponential	0.000	0.000
Weibull	0.000	0.000
Burr	0.003	0.000

Table 5
D-test results for price durations, Gaussian kernel

stock	in sample			out of sample		
	(0,0)	(2.5,2.5)	(30,1)	(0,0)	(2.5,2.5)	(30,1)
Boeing						
Exponential	0.000	0.000	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000	0.000	0.000
Burr	0.000	0.140	0.688	0.001	0.000	0.421
Coca-Cola						
Exponential	0.000	0.000	0.921	0.000	0.010	0.939
Weibull	0.000	0.000	0.949	0.000	0.070	0.901
Burr	0.001	0.032	0.985	0.041	0.337	0.709
Disney						
Exponential	0.000	0.000	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000	0.000	0.000
Burr	0.075	0.291	0.388	0.000	0.000	0.000
Exxon						
Exponential	0.000	0.000	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000	0.000	0.000
Burr	0.000	0.000	0.205	0.005	0.036	0.226
IBM						
Exponential	0.000	0.000	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000	0.000	0.000
Burr	0.000	0.000	0.099	0.000	0.000	0.003

The weighting scheme (x, y) is such that the first x and last y percent of the sample are trimmed out.

Table 6
H-test results for price durations, Gaussian kernel

stock	in sample (0,0)	in sample (5,20)	out of sample (0,0)	out of sample (5,20)
Boeing				
Exponential	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000
Burr	0.000	0.000	0.273	0.000
Coca-Cola				
Exponential	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000
Burr	0.000	0.000	0.000	0.000
Disney				
Exponential	0.055	0.000	0.000	0.023
Weibull	0.134	0.000	0.000	0.319
Burr	0.350	0.017	0.000	0.000
Exxon				
Exponential	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000
Burr	0.000	0.000	0.000	0.000
IBM				
Exponential	0.000	0.000	0.000	0.000
Weibull	0.000	0.000	0.000	0.000
Burr	0.000	0.000	0.000	0.000

The weighting scheme (x, y) is such that the first x and last y percent of the sample are trimmed out.

Table 7
The BDS test for serial independence, p-values

stock	in sample			out of sample		
	$m = 2$	$m = 3$	$m = 4$	$m = 2$	$m = 3$	$m = 4$
Boeing						
Exponential	0.000	0.000	0.004	0.019	0.012	0.039
Weibull	0.000	0.001	0.006	0.019	0.012	0.042
Burr	0.002	0.002	0.014	0.152	0.161	0.247
Coca-Cola						
Exponential	0.253	0.270	0.092	0.000	0.000	0.000
Weibull	0.252	0.269	0.093	0.000	0.000	0.000
Burr	0.253	0.272	0.099	0.000	0.000	0.000
Disney						
Exponential	0.242	0.230	0.163	0.241	0.100	0.135
Weibull	0.248	0.233	0.167	0.249	0.101	0.134
Burr	0.260	0.230	0.154	0.240	0.102	0.132
Exxon						
Exponential	0.053	0.039	0.073	0.114	0.053	0.019
Weibull	0.051	0.038	0.071	0.114	0.054	0.019
Burr	0.038	0.026	0.054	0.163	0.803	0.027
IBM						
Exponential	0.000	0.000	0.000	0.001	0.001	0.001
Weibull	0.000	0.000	0.000	0.001	0.001	0.001
Burr	0.000	0.000	0.001	0.002	0.001	0.002

The BDS test was computed using embedding dimension m and tuning parameter ϵ set to the standard deviation as recommended by Brock et al. (1996).