Nonparametric specification tests for conditional duration models

Marcelo Fernandes
Joachim Grammig
Nonparametric specification tests for conditional duration models

Marcelo Fernandes  Joachim Grammig
Getulio Vargas Foundation  University of Tübingen
mfernand@fgv.br  joachim.grammig@uni-tuebingen.de

Abstract: This paper deals with the testing of autoregressive conditional duration (ACD) models by gauging the distance between the parametric density and hazard rate functions implied by the duration process and their non-parametric estimates. We derive the asymptotic justification using the functional delta method for fixed and gamma kernels, and then investigate the finite-sample properties through Monte Carlo simulations. Although our tests display some size distortion, bootstrapping suffices to correct the size without compromising their excellent power. We show the practical usefulness of such testing procedures for the estimation of intraday volatility patterns.

JEL Classification: C14, C41, C52.

Keywords: duration models, gamma kernel, hazard rate, specification testing.

Acknowledgements: We are indebted to three anonymous referees, Renato Flóres, Søren Johansen, and many seminar participants for helpful comments, as well as to Sandra Vuletić and Bernardo de Sá Mota for valuable research assistance. Fernandes acknowledges with gratitude the financial support from CNPq-Brazil and a Jean Monnet Fellowship at the European University Institute. The usual disclaimer applies.

Corresponding author: Marcelo Fernandes, Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo 190, 22253-900 Rio de Janeiro, RJ, Brazil. Telephone: +55 21 2559 5827, Fax: +55 21 2553 8821.
1 Introduction

The availability of financial transactions data has raised the level of 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, which motivates the use of conditional duration models. Therefore, it is no wonder that microstructure studies employing conditional duration models abound in the literature (e.g. Lo, MacKinlay and Zhang, 2002). In particular, price durations (time between market maker’s mid-quote changes) are closely linked to the instantaneous volatility of the mid-quote price process (Engle and Russell, 1998) and thus they play an interesting role in option pricing (Prigent, Renault and Scaillet, 2001) and intraday risk management (Giot, 2000). Trade and volume durations mirror in turn features such as market liquidity and the information arrival rate (Gouriéroux, Jasiak and Le Fol, 1999).


Despite the boom in 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 residuals are independent and identically distributed (iid). All papers use the Ljung-Box statistic to test for serial correlation, yet only a few tests whether the
distribution of the error term (i.e. standardized duration) is correctly specified. Engle and Russell (1998) check the first and second moments of the residuals with particular attention to measuring excess dispersion, while others use QQ-plots (Bauwens and Veredas, 1999) and Bartlett identity tests (Prigent et al., 2001). Grammig and Wellner (2002) take a different approach by estimating and testing conditional duration models using a GMM framework. Bauwens, Giot, Grammig and Veredas (2000) employ the techniques developed by Diebold, Gunther and Tay (1998) to evaluate density forecasts of conditional duration models.

At first sight, misspecification of the distribution of the standardized duration may seem unimportant given that quasi maximum likelihood (QML) methods based on distributions belonging to the standard gamma family (with two parameters), such as the exponential, provide consistent estimates (Drost and Werker, 2003). However, QML estimation of conditional duration models may perform quite poorly in finite samples. Consider, for instance, a data generating process with a nonmonotonic baseline hazard rate function. Estimation by QML using the exponential distribution fails to produce sound results even in quite large samples (Grammig and Maurer, 2000). Further, the misspecification of the baseline distribution has quite serious implications for models that attempt to uncover the link between duration and volatility as in Ghysels and Jasiak (1998) and Engle (2000). Indeed, the success of option pricing and risk management procedures based on intraday volatility estimates from price duration models depends heavily on the appropriate specification of the baseline hazard rate function (Giot, 2000; Prigent et al., 2001).

This paper develops two-step procedures to test the distribution of the error term in a conditional duration model. The first step consists in estimating the conditional duration process by QML so as to obtain residuals that consistently estimate the errors. The second step then gauges the closeness between parametric and nonparametric estimates of the baseline density and hazard rate functions of the residuals. There is no novelty in the idea of comparing a consistent estimator under correct parameterization 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. Smoothing misspecification tests abound in the recent literature. See, for example, Kozek (1991), Wooldridge (1992), Fan and Gencay (1993), Härdle and Mammen (1993), Fan (1994), Hong and White
Our tests are not only simple, but also have some desirable properties. In contrast to Bartlett identity tests (Chesher, Dhaene, Gouriéroux and Scaillet, 1999), they examine the whole distribution of the 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 the errors (i.e. standardized durations) with estimated residuals. Monte Carlo simulations moreover indicate that some versions of our tests entail excellent performance in terms of finite-sample size and power.

The remainder of this paper is organized 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 behavior 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 of the asymptotic and bootstrap-based variants of our tests through Monte Carlo simulations. Section 6 applies our testing procedures to assess the performance of the linear ACD model for Exxon price durations at the New York Stock Exchange (NYSE), so as to highlight the effects of misspecifying the baseline distribution on the instantaneous volatility estimation. Section 7 summarizes the main results and offers concluding remarks. For ease of exposition, an appendix collects all proofs and technical lemmas.

2 Conditional duration models

Let $z_i = \psi_i \epsilon_i \geq 0$, where the duration $z_i = t_i - t_{i-1}$ denotes the time elapsed between events occurring at time $t_i$ and $t_{i-1}$, the conditional expected duration process $\psi_i = E(z_i | I_{i-1})$ is independent of the standardized duration $\epsilon_i$, 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 mean process

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

where $u_i | I_{i-1} \sim N(0, \sigma_u^2)$ and $\phi$ is a vector of parameters. For instance, Engle and Russell’s (1998) linear ACD(1,1) model defines $\psi_i = \omega + \alpha z_{i-1} + \beta \psi_{i-1}$,
whereas Fernandes and Grammig (2003) assume the more flexible specification
\[
\psi_i^\lambda = \omega + \alpha \psi_i^\lambda \left[ \left| \epsilon_{i-1} - b \right| + c (\epsilon_{i-1} - b) \right] + \beta \psi_i^\lambda.
\]

If the aim is to model microstructure effects, one can also incorporate additional predetermined variables (Engle and Russell, 1998; Bauwens and Giot, 2000).

Further, suppose that the standardized duration \( \epsilon_i \) is iid with Burr density
\[
f_B (\epsilon_i; \theta_B) = \frac{\kappa \xi_B^{\kappa} \epsilon_i^{\kappa-1}}{(1 + \gamma \xi_B \epsilon_i^\gamma)^{1+\gamma}},
\]  
where \( \theta_B = (\kappa, \gamma) \), \( \kappa > \gamma > 0 \) and
\[
\xi_B \equiv \frac{\Gamma(1 + 1/\kappa) \Gamma(1/\gamma - 1/\kappa)}{\gamma^1+1/\kappa \Gamma(1 + 1/\gamma)}.
\] (3)

It is readily seen that the conditional density of \( z_i \) is also Burr with parameter vector \( (\xi_B^{\kappa} \psi_i^{-\kappa}, \kappa, \gamma) \). Accordingly, the conditional hazard rate function reads
\[
H_B \left( z_i \bigg| I_{i-1}; \theta_B \right) = \frac{\kappa \xi_B^{\kappa} \psi_i^{-\kappa} z_i^{\kappa-1}}{1 + \gamma \xi_B^{\kappa} \psi_i^{-\kappa} z_i^\gamma},
\] (4)
which is nonmonotonic with respect to the standardized duration if \( \kappa > 1 \).

When \( \gamma \) shrinks to zero, (2) reduces to a Weibull distribution, viz.
\[
f_W (\epsilon_i; \theta_W) = \kappa \xi_W^{\kappa} \epsilon_i^{\kappa-1} \exp (-\xi_W \epsilon_i^\kappa),
\] (5)
where \( \theta_W = \kappa \) and \( \xi_W = \Gamma(1 + 1/\kappa) \). Accordingly, the conditional distribution of the duration process is also Weibull and the conditional hazard rate function reads \( H_W \left( z_i \bigg| I_{i-1}; \theta_W \right) = \kappa \xi_W^{\kappa} \psi_i^{-\kappa} z_i^{\kappa-1} \). In contrast to the Burr case, the conditional hazard rate implied by the Weibull distribution is monotonic. It decreases with the standardized duration for \( 0 < \kappa < 1 \), increases for \( \kappa > 1 \) and remains constant for \( \kappa = 1 \). In the latter case, the Weibull coincides with the exponential distribution and the conditional hazard rate function of the duration process is simply \( H_E \left( z_i \bigg| I_{i-1} \right) = \psi_i^{-1} \).

As an alternative, Lunde (1999) puts forward the generalized gamma ACD model, where \( \epsilon_i \) is iid with density
\[
f_G (\epsilon_i; \theta_G) = \frac{\xi_G^{\gamma \kappa} \kappa \epsilon_i^{\gamma-1}}{\Gamma(\gamma)} \exp (-\xi_G \epsilon_i^\gamma),
\] (6)
where \( \theta_G = (\kappa, \gamma) \) and \( \xi_G \equiv \Gamma(\gamma + 1/\kappa)/\Gamma(\gamma) \). The generalized gamma distribution nests the standard gamma (\( \kappa = 1 \)), exponential (\( \kappa = \gamma = 1 \)) and Weibull (\( \gamma = 1 \)) distributions, though it is nonnested with respect to the Burr distribution. The baseline hazard rate has no closed-form solution because it depends
on the incomplete gamma integral $I(\epsilon, \gamma) \equiv \int_0^{\epsilon} u^{\gamma-1} \exp(-u) \, du$. Nonetheless, it is possible to derive its shape properties according to the parameter values (Glaser, 1980). If $\kappa \gamma < 1$, the hazard rate is decreasing for $\kappa \leq 1$, and U-shaped for $\kappa > 1$. Conversely, if $\kappa \gamma > 1$, the hazard rate is increasing for $\kappa \geq 1$, and inverted U-shaped for $\kappa < 1$. Lastly, if $\kappa \gamma = 1$, the hazard rate is decreasing for $\kappa < 1$, constant for $\kappa = 1$ (exponential case), and increasing for $\kappa > 1$.

Albeit Engle and Russell (1998) suggest the use of exponential and Weibull distributions, the Burr and the generalized gamma ACD models seem to deliver better results for both price and trade durations (Lunde, 1999; Bauwens et al., 2000; Zhang et al., 2001). It remains the fact however that quasi maximum likelihood estimates are consistent only if based on the standard gamma distribution with two parameters (Drost and Werker, 2003).

3 Specification tests

As conditional duration models are usually estimated by the (quasi) maximum likelihood method, 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 $\chi^2$-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 usually quite slack, and are more likely to yield inconclusive results.

In the following, we design a testing strategy to check the parametric specification of the distribution of the standardized duration $\epsilon_i$ by matching density functionals. More precisely, we assume that the conditional mean process $\psi_i$ is correctly specified and then test whether there is any value $\theta_0$ of the parameter vector such that the true and parametric density functions of the standardized duration coincide almost everywhere. We first consider the general problem of testing the null

\[ H_0 : \exists \theta_0 \in \Theta \text{ such that } f(x, \theta_0) = f(x) \]  

against the alternative hypothesis that there is no such $\theta_0 \in \Theta$ for a nonnegative random variable $X$ with independent observations $x_1, \ldots, x_n$. In our context, we have a further twist because the standardized duration, which plays the
role of the nonnegative random variable \( X \), is unobservable. We thus extend the results in Section 4.4 to testing the estimated residuals of the conditional duration model \( \hat{\epsilon}_i = z_i/\hat{\psi}_i \) \((i = 1, \ldots, n)\).

The true cumulative distribution function \( F \) of the standardized durations and the density \( f \) are of course unknown, otherwise we could merely verify whether they belong to the specified parametric family of distributions. Accordingly, we estimate the density function using kernel methods, so as to have consistent estimates irrespective of the parametric specification of the distribution. Because the parametric density estimator is consistent only under the null, it is natural to develop tests that gauge the closeness between these two density estimates.

Our first testing procedure, which we label the D-test, rests on measuring the following distance
\[
\Phi_f = \int \mathbb{I}(x \in S) [f(x, \theta) - f(x)]^2 dF(x),
\]
where \( \int_x \) denotes the integral over the support of the density function of \( x \) and \( \mathbb{I}(\cdot) \) is the indicator function. The distance (8) weighs the difference between the parametric and nonparametric estimators according to their relevance as measured by the density function. Further, it is nonnegative with equality holding if and only if \( f(x) \) coincides with \( f(x, \theta) \) for every \( x \in S \). We introduce the subset \( S \) so as to avoid regions in which density estimation is unstable.

The sample analog of (8) reads
\[
\Phi_f = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_i \in S) \left[ f(x_i, \hat{\theta}) - \hat{f}(x_i) \right]^2,
\]
where \( \hat{\theta} \) and \( \hat{f}(\cdot) \) denote pointwise consistent estimates of the true parameter \( \theta_0 \) and density \( f(\cdot) \), respectively. The null hypothesis is then rejected if the D-test statistic \( \Phi_f \) is large enough. One could also work with other measures of closeness such as the integrated (rather than mean) square difference as in Bickel and Rosenblatt (1973) and Fan (1994).

By virtue of the one-to-one mapping linking hazard rate and density functions, the null hypothesis (7) implies that there exists a \( \theta_0 \in \Theta \) such that the hazard rate function implied by the parametric model \( H_{\theta_0}(\cdot) \) equals the true hazard function \( H_f(\cdot) \). Accordingly, we consider a second test, which we refer to as the H-test, relying on the statistic
\[
\Lambda_f = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_i \in S) \left[ H_{\hat{\theta}}(x_i) - \hat{H}_f(x_i) \right]^2,
\]
where $\mathcal{H}_g$ and $\mathcal{H}_f$ are the parametric and nonparametric estimates of the baseline hazard rate function, respectively. From the discussion above, (10) is close to zero under the null, while it is large under the alternative.

The next section shows that the limiting distribution of the D- and H-test statistics do not depend on how one estimates the parametric density function. This follows from the fact that the nonparametric density estimation converges at a slower rate than the parametric density estimation. One could therefore employ maximum likelihood methods. Alternatively, to provide a minimum-distance flavor to the D- and H-tests, one can estimate the parametric density by minimizing (9) and (10), respectively. The resulting estimators $\hat{\theta}_D \equiv \arg\min_{\theta \in \Theta} \Phi_f$ and $\hat{\theta}_H \equiv \arg\min_{\theta \in \Theta} \Lambda_f$ belong to the class of M-estimators discussed by Newey (1994) as they hinge on a two-step procedure in which the first step involves a kernel estimation and the second step solves a minimization problem.

4 Asymptotic justification

In what follows, we derive asymptotic results for the test statistics and their implied M-estimators. In fact, the limiting behavior of the D-test was already investigated by Aït-Sahalia (1996), who extends Bickel and Rosenblatt’s (1973) framework to estimate and test diffusion processes. Accordingly, the assumptions we impose are quite similar and the asymptotic results are the same up to a weighting scheme.

4.1 Assumptions

Let $x_1, \ldots, x_n$ denote independent observations on a random variable $X$ with probability density function $f(x), x \in [0, \infty)$. We consider the following set of regularity conditions.

**A1** The density function $f$ is continuously differentiable up to order $s + 1$ and its derivatives are bounded and square integrable. Further, $f$ is bounded away from zero on the compact interval $S$.

**A2** 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 $c_K \equiv \int_0^1 K^2(u) \, du$ and $v_K \equiv \int_0^1 \left[ \int_0^1 K(u)K(u + v) \, du \right]^2 \, dv$. 

8
A3 As the sample size $n$ grows, the bandwidths for the fixed and gamma kernels are such that $h_n = o\left( n^{-1/(2s+1)} \right)$ and $b_n = o\left( n^{-4/9} \right)$, respectively.

A4 The parameter space $\Theta \subset \mathbb{R}^k$ is compact. Let $\zeta(\cdot, \theta)$ denote the parametric density function $f(\cdot, \theta)$ for the D-test and the baseline hazard rate function $H(\cdot, \theta)$ for the H-test. In a neighborhood of the true parameter $\theta_0$, $\zeta(\cdot, \theta)$ is twice continuously differentiable in $\theta$ with uniformly bounded second-order partial derivatives and the matrix $E \left[ \frac{\partial}{\partial \theta} \zeta(\cdot, \theta) \frac{\partial}{\partial \theta} \zeta(\cdot, \theta) \right]$ has full rank.

A5 Consider $f_*$ and $f_+$ in a neighborhood $N_f$ of the true density $f$. Then, the leading term $\partial f$ that drives the asymptotic distribution of the implied M-estimators is such that

(i) $E|\partial f|^{3+r} < \infty$, for some $r > (3 + \eta)(3 + \eta/2)/\eta$ and $\eta > 0$

(ii) $E \sup_{f \in N_f} |\partial f_*|^2 < \infty$

(iii) $E|\partial f_* - \partial f_+|^2 \leq c\|f_* - f_+\|_{L(\infty, m)}$

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

Assumption A1 requires that the density function is smooth enough to admit a functional Taylor expansion. Although assumption A2 provides enough room for higher-order kernels, in what follows, we implicitly assume that the kernel is of second order ($s = 2$) as in Hall (1984). In particular, we will focus attention on Ghosh and Huang’s (1991) optimal uniform kernel for which $e_K = (2\sqrt{3})^{-1}$ and $v_K = (3\sqrt{3})^{-1}$. Assumption A3 induces undersmoothing in order to simplify the asymptotic bias of the test statistics. The optimal rate would imply additional bias terms as in Fan (1994). Assumption A4 guarantees that the functionals $\theta^D_f$ and $\theta^H_f$ implied by the M-estimators are well defined. Lastly, it follows from A5 that one can consistently estimate the asymptotic variance of the M-estimators using a nonparametric correction à la Newey and West (1987).

4.2 Matching the density function

The D-test gauges the discrepancy between the parametric and nonparametric estimates of the density function. It follows from (9) that the functional of interest is

$$
\Phi_f = \int_S I(x \in S) \left[ f(x, \theta_f) - \hat{f}(x) \right]^2 dF_n(x),
$$

(11)
where \( \theta_f \) is the functional implied by the estimator of \( \theta \) and \( F_n \) denotes the empirical cumulative distribution function. Assume further that it admits the following (von Mises) functional expansion

\[
\Phi_f = \Phi_f + D\Phi_f(h_x) + \frac{1}{2} D^2\Phi_f(h_x, h_x) + O\left(\|h_x\|^3\right), \tag{12}
\]

where \( h_x = h(x) = \bar{f}(x) - f(x) \) and \( \| \cdot \| \) denotes the \( L^2 \) norm. By the Riesz representation theorem, the functional derivative \( D\Phi_f(\cdot) \) has a dual representation of the form \( D\Phi_f(h_x) = \Psi_f(x)h_x \, dx \). It follows from Aït-Sahalia’s (1994) functional delta method that \( \Psi_f \) stands for the leading term that drives the asymptotic distribution of \( \Phi_f \). If the first functional derivative is degenerate, then the asymptotic distribution is driven by the second-order term of the expansion. See also von Mises (1947), Serfling (1980), and Ren and Sen (2001).

Let \( f_x \) and \( f_{x, \theta} \) denote the true and parametric density functions evaluated at \( x \), respectively. The first functional derivative of \( \Phi_f \) reads

\[
D\Phi_f(h_x) = \int_S (f_{x, \theta} - f_x)^2 h_x \, dx + 2 \int_S \left[ \frac{\partial f_{x, \theta}}{\partial \theta} D\theta_f(h_x) - h_x \right] (f_{x, \theta} - f_x) f_x \, dx, \tag{13}
\]

where \( D\theta_f(\cdot) \) denotes the first derivative of the functional \( \theta_f \) implied by the QML estimator. As \( D\Phi_f(h_x) \) is singular under the null, the limiting distribution of \( \Phi_f \) depends on the second functional derivative, namely

\[
D^2\Phi_f(h_x, h_x) = 2 \int_S \frac{\partial f(x, \theta_f)}{\partial \theta} \frac{\partial f(x, \theta_f)}{\partial \theta'} [D\theta_f(h_x)]^2 f_x \, dx
- 4 \int_S \frac{\partial f(x, \theta_f)}{\partial \theta} D\theta_f(h_x) f_x h_x \, dx
+ 2 \int_S f_x h_x^2 \, dx. \tag{14}
\]

However, the first and second terms of the right-hand side do not affect the asymptotic distribution of the test statistic. Aït-Sahalia (1994) shows indeed that the asymptotics are driven by the unsmoothest term of the first nondegenerate derivative because it converges at a slower rate. The third term contains a Dirac mass in its inner product representation, and thus leads the asymptotics.

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

\[
\hat{\tau}_n^D = \frac{nh_n^{1/2}\Phi_f - h_n^{-1/2}\hat{\sigma}_D}{\hat{\sigma}_D} \quad \overset{d}{\longrightarrow} N(0, 1),
\]

where \( \hat{\sigma}_D \) and \( \hat{\sigma}_D^2 \) are respectively consistent estimates of \( \sigma_D \equiv \epsilon_K E[\|x \in S\| f_x] \) and \( \sigma_D^2 \equiv \nu_K E[\|x \in S\| f_x^3] \).

**Proof.** See Aït-Sahalia (1996).
To obtain consistent estimates of $\delta_D$ and $\sigma^2_D$, one may use the empirical distribution to compute the expectation and plug in the corresponding kernel density estimate, then yielding

$$\hat{\delta}_D = e_K \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_i \in S) \hat{f}(x_i)$$

$$\hat{\sigma}^2_D = v_K \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_i \in S) \left[ \hat{f}(x_i) \right]^3.$$ 

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, the test statistic $\hat{\tau}_D$ may perform poorly due to the boundary bias that haunts nonparametric 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 $f_Y(y) = f_X[\exp(y)] \exp(y)$. An alternative rests on estimating the density function using asymmetric kernels to benefit from the fact that they never assign weight outside the density support (see Bouezmarni and Rolin, 2001, Bouezmarni and Scaillet, 2002, Chen, 1999 and 2000, Scaillet, 2003). In particular, the gamma kernel

$$K_{x/b_n + 1, b_n}(u) = \frac{u^{x/b_n} \exp(-u/b_n)}{\Gamma(x/b_n + 1)b_n^{x/b_n + 1}} I(u \geq 0)$$

with bandwidth $b_n$ provides estimates that are consistent, nonnegative, boundary bias free, and achieve the optimal rate of convergence for the mean integrated error for any density function whose support is bounded from the origin (Bouezmarni and Scaillet, 2002; Chen, 2000). We therefore consider a second version of the D-test in which the density estimation uses a gamma kernel.

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

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

where $\hat{\delta}_G$ and $\hat{\sigma}_G^2$ are consistent estimates of $\delta_G = \frac{1}{2\sqrt{\pi}} E \left[ [I(x \in S)x^{-1/2}f_x] \right]$ and $\sigma_G^2 = \frac{1}{2}\sqrt{\pi} E \left[ [I(x \in S)x^{-1/2}f_x^2] \right]$, respectively.

As above, consistent estimates of $\delta_G$ and $\sigma_G^2$ are readily available by plugging in the empirical distribution and gamma kernel density estimate $\hat{f}$, viz.

$$\hat{\delta}_G = \frac{1}{2\sqrt{\pi}n} \sum_{i=1}^{n} \mathbb{I}(x_i \in S)x_i^{-1/2} \hat{f}(x_i)$$
\[ \tilde{\sigma}_G^2 = \frac{1}{2\pi n} \sum_{i=1}^{n} I(x_i \in S)x_i^{-1/2}[\tilde{f}(x_i)]^3. \]

Consider now, as in Aït-Sahalia, Bickel and Stoker (2001), the sequence of local alternatives

\[ H_{1n}^D : \sup_{x \in S} \big| f^{[n]}(x, \theta) - f^{[n]}(x) - \varepsilon_n \ell_D(x) \big| = o(\varepsilon_n), \] (18)

where \( \|f^{[n]} - f\| = o(\varepsilon_n) \) and

\[ \varepsilon_n = \begin{cases} n^{-1/2}h_n^{-1/4} & \text{when using a fixed kernel} \\ n^{-1/2}h_n^{-1/8} & \text{when using a gamma kernel.} \end{cases} \] (19)

Assume further that \( \ell_D(x) \) is such that \( \ell_S^D \equiv E[I(x \in S)\ell_D^2(x)] \) exists and \( E[\ell_D(x)] = 0. \) The next result illustrates the fact that both versions of the D-test have nontrivial power under local alternatives that shrink to the null at the corresponding rate \( \varepsilon_n. \)

**Proposition 3.** Under the sequence of local alternatives \( H_{1n}^D \) and assumptions A1 to A4, \( \tilde{\tau}_D^n d \rightarrow N(\ell_S^D/\sigma_D, 1), \) whereas \( \hat{\tau}_D^n d \rightarrow N(\ell_S^D/\sigma_G, 1). \)

To maximize the power of both versions of the D-test, one could consider the most favorable scenario to the parametric model by utilizing the M-estimator \( \hat{\theta}_D^n. \) The corresponding implicit functional then is

\[ \int_S \frac{\partial f(x, \theta_D^n)}{\partial \theta} \left[ f(x, \theta_D^n) - f(x) \right] f(x) \, dx \equiv 0, \] (20)

which produces

\[ D\theta_D^n(h_x) = \left\{ \int_S \frac{\partial f(x, \theta)}{\partial \theta} \frac{\partial f(x, \theta)}{\partial \theta'} \frac{\partial f(x)}{\partial \theta} \, dx \right\}^{-1} \int_S \frac{\partial f(x, \theta)}{\partial \theta} f(x) h(x) \, dx. \] (21)

Accordingly, the limiting distribution is driven by

\[ \vartheta_f^D(x) = I(x \in S) \left\{ \int_S \frac{\partial f(x, \theta)}{\partial \theta} \frac{\partial f(x, \theta)}{\partial \theta'} \frac{\partial f(x)}{\partial \theta} \, dx \right\}^{-1} \frac{\partial f(x, \theta)}{\partial \theta} f(x). \] (22)

**Proposition 4.** Under the null and assumptions A1 to A4, \( n^{1/2}(\hat{\theta}_D^n - \theta_0) \) \( d \rightarrow N(0, \Omega_D), \) where \( \Omega_D \equiv \sum_{h=0}^{\infty} \text{Cov} \left[ \vartheta_f^D(x_i), \vartheta_f^D(x_{i+h}) \right] \) is the long run covariance matrix of \( \vartheta_f^D. \) In addition, if assumption A5 holds, it suffices to plug \( \hat{\theta}_D^n \) into \( \vartheta_f^D \) 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 nonparametric estimates of the baseline hazard rate. The motivation is simple. The usual densities associated with duration models may engender fairly similar shapes depending on the parameter values. In turn, they hatch very different hazard rate functions: flat for the exponential, monotonic for the Weibull and nonmonotonic for the Burr and the generalized gamma distributions.

From (10), the functional of interest reads

$$\Lambda_f = \int_0^\infty \mathbb{1}(x \in S) [H_\theta(x) - H_f(x)]^2 dF_n(x)$$

$$= \int_S [H_\theta(x) - H_f(x)]^2 dF_n(x), \quad (23)$$

Suppose that (23) admits a second-order Taylor expansion about the true density, viz.

$$\Lambda_f = \Lambda_f + D\Lambda_f(h_x) + \frac{1}{2} D^2\Lambda_f(h_x, h_x) + O(\|h_x\|^3), \quad (24)$$

where $\Lambda_f = \int_S [H_\theta(x) - H_f(x)]^2 f(x) dx$ and $h_x = h(x) = f(x) - f(x)$ as before. The first functional derivative is then

$$D\Lambda_f(h_x) = \int_S [H_\theta(x) - H_f(x)]^2 h_x dx$$

$$+ 2 \int_S [H_\theta(x) - H_f(x)] \left[ \frac{\partial H_\theta(x)}{\partial \theta} D\theta f(h_x) - D\mathcal{H}_f(h_x) \right] dF(x), \quad (25)$$

where

$$D\mathcal{H}_f(h_x) = \frac{h(x) - h_f(x)}{S_x} \int_u \mathbb{1}(u < x) h(u) du \quad (26)$$

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

$$D^2\Lambda_f(h_x, h_x) = 2 \int_S [D\mathcal{H}_f(h_x)]^2 dF(x)$$

$$+ 2 \int_S \frac{\partial H_\theta(x)}{\partial \theta} \frac{\partial H_\theta(x)}{\partial \theta'} [D\theta f(h_x)]^2 dF(x)$$

$$- 4 \int_S \frac{\partial H_\theta(x)}{\partial \theta} D\theta f(h_x) D\mathcal{H}_f(h_x) dF(x). \quad (27)$$

It turns out that the first term leads the asymptotics as it contains the un-smoothest term of the expansion.
Proposition 5. Under the null and assumptions A1 to A4, the statistic
\[ \tilde{\tau}_n^H = \frac{nh_n^{1/2} \Lambda_f - h_n^{-1/2} \hat{\lambda}_H}{\tilde{\varsigma}_H} \xrightarrow{d} N(0, 1), \tag{28} \]
where \( \hat{\lambda}_H \) and \( \tilde{\varsigma}_H^2 \) are consistent estimates of \( \lambda_H = \epsilon_K E[1(x \in S)\mathcal{H}_f(x)/S_x] \) and \( \varsigma_H^2 = \nu_K E[1(x \in S)\mathcal{H}_f^2(x)/S_x] \), respectively.

To obtain consistent estimates of \( \lambda_H \) and \( \varsigma_H^2 \), one may compute the expectation using the empirical distribution \( F_n \) and plug in the corresponding kernel estimates \( \hat{f} \) and \( \hat{F} \) of the density and cumulative distribution functions to estimate the hazard rate and survival functions. This yields
\[ \hat{\lambda}_H = \epsilon_K \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \in S) \hat{f}(x_i) [1 - \hat{F}(x_i)]^{-2} \]
\[ \tilde{\varsigma}_H^2 = \nu_K \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \in S) [\hat{f}(x_i)]^3 [1 - \hat{F}(x_i)]^{-4}. \]

Other nonparametric estimators for the hazard rate and survival functions could also be used. See Ramlau-Hansen (1983), Nielsen (1998), and Nielsen and Tanggaard (2001).

In contrast to the density function, in general, there is no closed form solution for the hazard rate of the log-standardized 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.

Proposition 6. Under the null and assumptions A1 to A4, the statistic
\[ \tilde{\tau}_n^H = \frac{nb_n^{1/4} \Lambda_f - b_n^{-1/4} \hat{\lambda}_G}{\tilde{\varsigma}_G} \xrightarrow{d} N(0, 1), \tag{29} \]
where \( \hat{\lambda}_G \) and \( \tilde{\varsigma}_G^2 \) estimate consistently \( \lambda_G = \frac{1}{2\sqrt{\pi}} E\left[1(x \in S)x^{-1/2}\mathcal{H}_f(x)/S_x\right] \) and \( \varsigma_G^2 = \frac{1}{\sqrt{2\pi}} E\left[1(x \in S)x^{-1/2}\mathcal{H}_f^2(x)/S_x\right] \), respectively.

Again, consistent estimates of \( \lambda_G \) and \( \varsigma_G^2 \) follows by plugging in the gamma kernel density and cumulative distribution estimates \( \hat{f} \) and \( \hat{F} \), giving way to
\[ \hat{\lambda}_G = \frac{1}{2\sqrt{\pi} n} \sum_{i=1}^n \mathbb{I}(x_i \in S)x_i^{-1/2} \hat{f}(x_i) [1 - \hat{F}(x_i)]^{-2} \]
\[ \tilde{\varsigma}_G^2 = \frac{1}{\sqrt{2\pi} n} \sum_{i=1}^n \mathbb{I}(x_i \in S)x_i^{-1/2} [\hat{f}(x_i)]^3 [1 - \hat{F}(x_i)]^{-4}. \]

Consider next the following sequence of local alternatives
\[ H_{1n}^H : \sup_{x \in S} \left| \mathcal{H}_f^{[n]}(x, \theta) - \mathcal{H}_f^{[n]}(x) - \varepsilon_n \ell_H(x) \right| = o(\varepsilon_n), \tag{30} \]
where \( \|H_f^{[\alpha]} - H_f\| = o(\varepsilon_n^2) \) and \( \ell_H(x) \) is such that \( \ell_H^2 \equiv E[\mathbb{I}(x \in S)\ell_H(x)] \) exists and \( E[\ell_H(x)] = 0 \). It then follows that both versions of the H-test can distinguish alternatives that get closer to the null at the corresponding rate \( \varepsilon_n \) in (18) while keeping the power constant.

**Proposition 7.** Under the sequence of local alternatives \( H_{1n}^H \) and assumptions A1 to A4, \( \tau_n^H \overset{d}{\longrightarrow} N(\ell_{H}^S/\varsigma_H, 1) \), whereas \( \tau_n^H \overset{d}{\longrightarrow} N(\ell_{H}^S/\varsigma_G, 1) \).

Finally, consider the M-estimator \( \hat{\theta}_n^H \) that minimizes the distance between the parametric and nonparametric estimates of the baseline hazard rate function. The corresponding implicit functional is

\[
\int_S \frac{\partial H(x, \theta_n^H)}{\partial \theta} \left[ \mathcal{H}(x, \theta_n^H) - H_f(x) \right] dF(x) \equiv 0,
\]

which results in the following first functional derivative

\[
D\theta_n^H(h_x) = \left\{ \int_S \frac{\partial H(x, \theta)}{\partial \theta} \frac{\partial H(x, \theta)}{\partial \theta} dF(x) \right\}^{-1} \int_S \frac{\partial H(x, \theta)}{\partial \theta} D\mathcal{H}_f(h_x) dF(x).
\]

From (26), it is readily seen that

\[
\dot{\mathcal{H}}_f(x) = \mathbb{I}(x \in S) \left\{ \int_S \frac{\partial H(x, \theta)}{\partial \theta} \frac{\partial H(x, \theta)}{\partial \theta} dF(x) \right\}^{-1} \frac{\partial H(x, \theta)}{\partial \theta} \mathcal{H}_f(x)
\]

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

**Proposition 8.** Under the null and assumptions A1 to A4, \( n^{1/2}(\hat{\theta}_n^H - \theta_0) \overset{d}{\longrightarrow} N(0, \Omega_H) \), where \( \Omega_H \equiv \sum_{k=-\infty}^{\infty} \text{Cov} \left[ \theta_n^H(x_i), \theta_n^H(x_{i+k}) \right] \) is the long run covariance matrix of \( \theta_n^H \). In addition, provided that assumption A5 holds, it suffices to plug \( \hat{\theta}_n^H \) into \( \dot{\theta}_n^H \) and truncate the infinite sum as in Newey and West (1987) to obtain a consistent estimator of the asymptotic variance.

### 4.4 Nuisance parameter freeness

All results so far consider testing the distributional properties of a random variable \( X \) with independent observations \( x_1, \ldots, x_n \). In the context of conditional duration models, the interest is in testing the standardized durations \( \epsilon_i = z_i/\psi_i \), \( i = 1, \ldots, n \). However, \( \{\epsilon_i\} \) is unobservable and the testing procedure must then proceed using the estimated residuals \( \hat{\epsilon}_i = z_i/\hat{\psi}_i \), \( i = 1, \ldots, n \). In the following, we derive the conditions under which the D- and H-tests are nuisance parameter free, and hence there is no asymptotic cost in substituting residuals for errors.

To simplify notation, let \( e_i = e_i(\phi_0) = z_i/\psi_i(\phi_0) \) and \( \hat{e}_i = e_i(\hat{\phi}) = z_i/\psi_i(\hat{\phi}) \), where \( \hat{\phi} \) is a root–n consistent estimate of the true parameter vector \( \phi_0 \) in (1).
Engle and Russell (1998) show that this last assumption holds for the QML estimator based on the exponential distribution. We hereinafter assume that the regularity conditions ensuring the consistency and asymptotic normality of the QML estimation hold, including the correct specification of the conditional mean process given by (1). See Gallant and White (1988) and Lee and Hansen (1994) for in-depth discussions of those conditions. Alternatively, one could either employ semiparametric techniques as in Drost and Werker (2003) or follow a GMM estimation strategy as in Grammig and Wellner (2002).

Proposition 9. Under assumptions A1 to A4, the D- and H-tests are nuisance parameter free in that there is no asymptotic cost in substituting the root−n consistent estimates \( \hat{\epsilon}_i \)'s for the unobserved \( \epsilon_i \)'s.

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 nonparametric tests entail substantial size distortions in finite samples. For instance, Fan and Linton (2003) demonstrate how neglecting higher order terms that are close in order to the dominant term may provoke such distortions. Not surprisingly, we show that size distortions are indeed material, so that we also consider bootstrap-based versions of our tests as in Fan (1995) and Li and Wang (1998). As singled out by Horowitz and Savin (2000), bootstrapping permits the computation of meaningful size-corrected critical values, putting power figures into perspective. The results show that applying the standard bootstrap procedure to the estimated residuals (Horowitz, 2001, Section 4.1.1) works pretty well, eliminating size distortions without compromising the power of the tests.

We generate 15,000 realizations of a linear ACD(1,1) model, i.e.

\[
\psi_i = \omega + \alpha z_{i-1} + \beta \psi_{i-1},
\]

by drawing the errors \( \epsilon_i = z_i/\psi_i \) from four distributions: exponential, Weibull with \( \kappa = 0.6 \), Burr with \( \kappa = 1.3 \) and \( \gamma = 0.4 \), and the generalized gamma with \( \kappa = 0.5 \) and \( \gamma = 3 \). We set \( \alpha = 0.1 \) and \( \beta = 0.7 \) to match the typical estimates found in empirical applications. Further, we normalize the unconditional expected duration to one by imposing \( \omega = 1 - (\alpha + \beta) \) and then set \( \psi_0 = 1 \) to initialize (34). Along with the full sample \( (n = 15,000) \), we also consider a subsample formed by the last 3,000 realizations so as to mitigate initial ef-
fects. These are typical sample sizes for data on trade and price durations, respectively. All results are based on 1,000 Monte Carlo replications.

For each replication and data generating process, we first compute QML estimates using the exponential distribution. The optimization procedure takes advantage of Han’s (1977) sequential quadratic programming algorithm, so as to accommodate general inequality constraints. Next, we examine the outcomes of our five tests: the D- and H-tests with Ghosh and Huang’s (1991) optimal uniform kernel and Chen’s (2000) gamma kernel applied to the residuals and the D-test with the optimal uniform kernel applied to log-residuals. We select the bandwidths by adjusting Silverman’s (1986) rule of thumb. The normal distribution serves as reference only for the log-standardized durations, the reference being the exponential otherwise. Bearing assumption A4 in mind, we also divide the rule-of-thumb bandwidths by $\log n$. For instance, the resulting bandwidth for the gamma kernel is

$$
\hat{b}_n = \frac{1}{\log n} \left( \hat{\lambda}/4 \right)^{-1/5} \left( 2 - \hat{\lambda} \right)^{-4/5} n^{-4/9},
$$

where $\hat{\lambda}$ is some consistent estimator of the exponential parameter $\lambda$, e.g., the inverse of the sample mean.

The frequency of rejection of the null hypothesis is then computed in order to evaluate the size and power of the six 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. 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 specifies an exponential distribution, whereas the true density is a Burr.

Figures 1 to 4 display the main results for $n = 3,000$ using Davidson and MacKinnon’s (1998) graphical representation to illustrate the finite-sample properties of the asymptotic and bootstrap-based tests for the ACD process with exponential, Weibull, Burr and generalized gamma errors, respectively. Each figure consists of two columns of charts. The first column corresponds to the asymptotic tests, whereas the second column illustrates the results for the bootstrap-based tests with 499 artificial samples. For every chart, the horizontal axis represents the significance level, while the vertical axis represents the

---

1. To conserve on space, we refrain from displaying similar graphs for the full sample ($n = 15,000$) given that, on balance, the results bear great resemblance. In particular, size distortions remain roughly constant, whereas the power improves mildly in general.
probability of rejection at that 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.

The asymptotic tests based on the optimal uniform kernel exhibit substantial size distortions unless applied to the log-residuals for which it is slightly liberal irrespective of the data generating process. Bootstrapping the estimated residuals however suffices to correct the size of the tests without affecting their power. In contrast, the asymptotic gamma-based tests perform relatively well both in terms of size and power, hence bootstrapping is not really necessary. As is apparent, our tests exhibit excellent power unless one wishes to distinguish between the generalized gamma and Burr distributions. The H-test with optimal kernel has the worst performance: It almost never rejects the Burr ACD specification when the true baseline distribution is the generalized gamma.

For the sake of comparison, we also run the overdispersion test as advocated by Engle and Russell (1998) for the ACD models with the exponential and Weibull distributions. The performance of the overdispersion test is quite disappointing. Figures 1 and 2 show that, despite its parametric nature, size distortions are material as also evinced by Chesher et al. (1999). Although bootstrapping successfully corrects the size of the overdispersion test, Figures 3 and 4 make evident that its power performance is surprisingly poor relative to the nonparametric tests. Altogether, the D- and H-tests with gamma kernel seem to entail the best overall performance, though the D-test for log-standardized durations is a fair alternative, especially if one wishes to avoid bootstrapping.

6 Empirical application

In this section, we use NYSE price duration data to assess the performance of the linear ACD model (34) with exponential, Weibull, Burr, and generalized gamma distributions employing the testing framework outlined in Section 3. Exxon price durations from NYSE’s Trade and Quote (TAQ) database were kindly provided by Luc Bauwens and Pierre Giot, who thoroughly describe the data in Bauwens and Giot (1998 and 2000) and Giot (2000).

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. 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 (2000). Durations between events recorded outside the regular opening hours of the NYSE as well as overnight spells are removed.

As documented by Giot (2000), 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. We thus consider seasonally adjusted price durations $z_i = Z_i/\varrho(t_i)$, where $Z_i$ is the raw price duration in seconds and $\varrho(\cdot)$ denotes an intraday seasonal factor. As in Engle and Russell (1998), we estimate the latter by averaging durations over thirty-minute intervals for each day of the week and fitting a cubic spline with nodes at each half hour. The resulting (seasonally adjusted) price durations $z_i$ then serve as input for the analysis.

The sample ranges from September to November 1996, consisting of 2,717 price durations. We use the first two thirds of the sample for estimation purposes, and reserve the remainder of the data to out-of-sample analysis. The descriptive statistics reported in Table 1 evince that Exxon price durations exhibit the two features that motivate ACD modeling: highly significant serial correlation and overdispersion.

6.1 Estimation results

We invoke (quasi) maximum likelihood methods to estimate the linear ACD model (34) with exponential, Weibull, Burr, and generalized gamma distributions. More precisely, as advanced in Section 5, we first compute the QML estimates using the exponential distribution. This yields

$$
\psi_i = 0.065 + 0.046 z_{i-1} + 0.890 \psi_{i-1},
$$

$$
(0.037) (0.016) (0.048)
$$

(35)

where figures in parentheses refer to QML standard errors. We next obtain the series of residuals and one-step ahead forecast errors by forming both in- and out-of-sample standardized durations $z_i/\hat{\psi}_i$.

Under the correct specification of the conditional mean, the residuals and forecast errors from (35) are both serially independent. We therefore check

$^2$ Alternatively, one could employ Veredas, Rodriguez-Poo and Espasa’s (2001) semiparametric approach to simultaneously estimate the seasonal component $\varrho(\cdot)$ and the parameters of the ACD model by local quasi-maximum likelihood.
whether the in- and out-of-sample residuals are serially independent by the means of 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), which is quite convenient given that we are testing estimated residuals rather than true errors. A simple log-transformation renders the linear ACD model additive, hence we work with the log-standardized durations. The results indicate that we cannot reject the null hypothesis of serial independence of the residuals and forecast errors at the 5% level of significance.\footnote{The complete set of results for the BDS test statistic with different embedding dimensions and tuning parameters are available from the authors upon request. The same applies for Ljung-Box test whose results indicate no rejection up to the 60th order.}

The linear ACD specification thus seems to do a pretty good job of accounting for the serial dependence of the Exxon price durations.

As the findings indicate that (35) is a congruent model, we can use the residuals to estimate by maximum likelihood the parameters of the Weibull, Burr and generalized gamma distributions. We find the following pointwise estimates (with their respective standard errors in parentheses): \( \hat{\kappa} = 0.962 \) (0.016) for the Weibull, \( \hat{\kappa} = 1.239 \) (0.042) and \( \hat{\gamma} = 0.432 \) (0.061) for the Burr, and \( \hat{\kappa} = 0.318 \) (0.056) and \( \hat{\gamma} = 8.052 \) (2.733) for the generalized gamma distribution. An informal likelihood comparison based on the usual information criteria rejects both the exponential and the Weibull ACD models, thus justifying the flexibility brought about by the Burr and generalized gamma formulations.

### 6.2 Test results

To further evaluate the performance of the four linear ACD specifications, we employ the testing procedures put forth in Section 3 to their estimated residuals and one-step ahead forecast errors. Although we also report the p-values for the asymptotic D-test with optimal uniform kernel for the log-residuals, we focus attention to the bootstrap versions of our tests. We refrain from reporting the results of the H-test with optimal uniform kernel because, as seen in Section 5, it entails a very poor small-sample performance (even after bootstrapping). Finally, as advocated by Engle and Russell (1998), we also use the asymptotic and bootstrap-based overdispersion tests to check the exponential and Weibull specifications. As in Section 5, all bootstrap variants of the tests consider 499 artificial samples to build the empirical distribution of the test statistic.
Table 2 reports the in- and out-of-sample results, which grant support only to the linear ACD model with the generalized gamma distribution. The other ACD specifications are less successful. The bootstrap-based tests indeed reject the exponential and Weibull ACD models in all instances. In contrast to the asymptotic D-test for log-standardized durations, whose results conform with the evidence given by the bootstrap-based tests, the asymptotic overdispersion test fails to reject the exponential and Weibull specifications for the forecast errors. This illustrates how misleading the overdispersion test can be if one simply ignores the fact that, in small samples, it features size distortions.

Interestingly, we find mixed results for the Burr ACD model both in- and out-of-sample. As opposed to the D- and H-tests with gamma kernel, the asymptotic and bootstrap-based variants of the D-test for log-standardized durations are unable to reject the Burr distribution for the residuals and forecast errors at the 5% level of significance. This result is likely to exemplify the lack of power of the D-test for log-standardized durations to distinguish the generalized gamma from the Burr distribution.

### 6.3 Intraday volatility estimates

To better illustrate the consequences of misspecifying the innovation distribution, we estimate the instantaneous volatility implied by our four specifications along the lines of Engle and Russell (1998). The instantaneous volatility $\sigma(t)$ is such that

$$
\sigma^2(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E \left[ \frac{P(t - \Delta t) - P(t)}{P(t)} \right]^2.
$$

As we are interested in price durations, we can hold the price change $c$ constant and then take the limit in (36) to estimate the instantaneous volatility by means of the linear ACD model. More specifically,

$$
\sigma^2(t) = \left( \frac{c}{P(t)} \right)^2 \mathcal{H}(t) \phi(t).
$$

where $\mathcal{H}(\cdot)$ denotes the conditional hazard function implied by the ACD specification and $\phi(t)$ corresponds to the seasonal component. Because the hazard function is defined for all $t > t_0$, it is possible to draw a continuous picture of the volatility over the trading day. As in Engle and Russell (1998), we compute annualized instantaneous volatilities by multiplying $\sigma^2(t)$ by the annual trading time in seconds.\(^4\)

\(^4\) There are 252 trading days per year, 6.5 hours per day, and 3,600 seconds per hour.
Figure 5 displays the instantaneous volatility estimates implied by each of the four ACD specifications for a typical trading day (Thursday, September 9, 1999). As the model selection procedure indicates that the generalized gamma ACD model is the best specification, the charts take its volatility estimates as reference in order to highlight the differences between the volatility estimates given by the other linear ACD models. As in Engle and Russell (1998), the spikes indicate event arrival times (quote changes), and the overall U-shaped volatility pattern over the course of the day confirms the previous findings in the intraday volatility literature (see, for example, Andersen and Bollerslev, 1997).

The variation of the instantaneous volatility implied by the exponential ACD model after the quote event in Figure 5 does not contradict the fact that the hazard rate function of the exponential distribution is flat. This is just a consequence of the seasonality of the quote intensity process. As \( \hat{\kappa} \) is slightly smaller than one, the instantaneous volatility implied by the Weibull ACD model is very similar to those implied by the exponential specification, though it monotonically decreases after a quote update. In contrast, the Burr and generalized gamma parameter estimates are such that their respective hazard function rates are nonmonotonic. Therefore, as implied by (37), it is not surprising that their implied instantaneous volatility estimates vary nonmonotonically after quote updates. Indeed, after a quote event, the instantaneous volatility seems to first decrease substantially, and then to sharply increase. It is worth noting that the generalized gamma specification predicts a much steeper slope than the Burr.

To sum up, the specifications with Weibull, Burr and generalized gamma baseline distributions all predict that it is most likely to observe a price event immediately after another price event. However, the intraday volatility estimates implied by the Weibull notably differ from those implied by the Burr and generalized gamma specifications. The instantaneous volatility decreases much faster for the Burr and generalized gamma than for the Weibull ACD model. This sort of instantaneous volatility pattern is consistent with the theoretical implications of Admati and Pfleiderer’s (1988) and Easley and O’Hara’s (1992) market microstructure models.

7 Concluding remarks

We propose two testing strategies that rely on gauging the discrepancy between parametric and nonparametric estimates of either the baseline density or haz-
ard rate function of standardized durations. Asymptotic theory is derived for nonparametric density estimation using both fixed and gamma kernels. The motivation for the latter is to avoid the boundary bias that plagues fixed kernel density estimation. All in all, our tests have attractive theoretical properties. They not only examine the whole distribution of the residuals instead of a limited number of moment restrictions, but they are also nuisance parameter free. Monte Carlo experiments show that the bootstrap-based versions of our tests have excellent performance in finite samples.

An empirical application of our tests shows that the linear ACD model with exponential, Weibull and Burr baseline distributions do not seem adequate to modeling Exxon price durations, as opposed to Lunde’s (1999) generalized gamma specification. The latter allows for a less restrictive hazard rate function, ensuring enough flexibility to estimate the instantaneous volatility of the mid-quote price process by means of conditional duration models.

Although this paper deals with specification tests for conditional duration models, there is no impediment to using such tests in other contexts. For instance, one could test GARCH-type models by checking whether the distribution of the errors is correctly specified. Similarly, Cox’s (1955) proportional hazard model implies testable restrictions in the hazard rate function.
Appendix: Proofs

Lemma 1. Consider the functional \( I_G = \int_x \varphi(x) \left( \hat{f}(x) - f(x) \right)^2 \, dx \), where \( \hat{f}(\cdot) \) is a pointwise gamma kernel estimate of \( f(\cdot) \). Under assumptions A1 and A3,
\[
n b_n^{1/4} I_G \xrightarrow{d} N \left( 0, \frac{1}{2 \pi} E \left[ x^{-1/2} \varphi(x) f(x) \right] \right),
\]
provided that the above expectations exist.


Lemma 2. Suppose that a functional \( \Upsilon_f \) is Fréchet differentiable relative to the Sobolev norm of order \((2, m)\) at the true density \( f \) with a regular functional derivative \( v_f \). Then, under assumptions A1 to A4, \( n^{1/2}(\Upsilon_f - \Upsilon_f) \xrightarrow{d} N(0, V) \), where \( V \equiv \sum_{k=\infty}^{\infty} \text{Cov}[v_f(x_i), v_f(x_{i+h})] \) is the long run covariance matrix of the functional derivative \( v_f \).

Proof. See Ait-Sahalia (1994).

Lemma 3. Suppose that the U-statistic \( U_n \equiv \sum_{1 \leq i < j \leq n} H_n(X_i, X_j) \) with symmetric variable function \( H_n(\cdot, \cdot) \) is centered and degenerate. If
\[
\frac{E_{X_1, X_2} \left\{ E_{X_1} \left[ H_n(X_1, X_1) H_n(X_1, X_2) \right] \right\} + \frac{1}{2} E_{X_1, X_2} \left[ H_n^2(X_1, X_2) \right]}{E_{X_1, X_2} \left[ H_n^2(X_1, X_2) \right]} \xrightarrow{d} 0
\]
as sample size grows, then
\[
U_n \xrightarrow{d} N \left( 0, \frac{n^2}{2} E_{X_1, X_2} \left[ H_n^2(X_1, X_2) \right] \right).
\]


Lemma 4. Consider the functional \( I_n = \int_x \varphi(x) \left( \hat{f}(x) - f(x) \right)^2 \, dx \), where \( \hat{f}(\cdot) \) is the pointwise fixed kernel estimate of \( f(\cdot) \). Assume further that \( \varphi(\cdot) \) is continuously differentiable and its first derivative is bounded and square integrable. Then, under assumptions A1 to A3,
\[
n h_n^{1/2} I_n - h_n^{-1/2} \epsilon_K E[\varphi(x)] \xrightarrow{d} N \left( 0, v_K E[\varphi^2(x)f(x)] \right),
\]
provided that the above expectations are finite.

Proof. Let \( r_n(x, X) = \varphi(x)^{1/2} K_{h_n}(x - X) \), where \( K_{h_n}(u) = h_n^{-1} K(u/h_n) \), and \( \bar{r}_n(x, X) = r_n(x, X) - E_X[r_n(x, X)] \). Consider then the following decomposition
\[
I_n = \int_x \varphi(x) \left( \hat{f}(x) - E\hat{f}(x) \right)^2 \, dx + \int_x \varphi(x) \left[ E\hat{f}(x) - f(x) \right]^2 \, dx \\
+ 2 \int_x \varphi(x) \left( \hat{f}(x) - E\hat{f}(x) \right) \left[ E\hat{f}(x) - f(x) \right] \, dx,
\]
24
or equivalently, \( I_n = I_{1n} + I_{2n} + I_{3n} + I_{4n} \), where

\[
I_{1n} = \frac{2}{n^2} \sum_{i<j} \int x \hat{r}_n(x, X_i) \hat{r}_n(x, X_j) \, dx
\]

\[
I_{2n} = \frac{1}{n^2} \sum_i \int x \hat{r}_n^2(x, X_i) \, dx
\]

\[
I_{3n} = \int \varphi(x) \left[ E\hat{f}(x) - f(x) \right]^2 \, dx
\]

\[
I_{4n} = 2 \int \varphi(x) \left[ \hat{f}(x) - E\hat{f}(x) \right] \left[ E\hat{f}(x) - f(x) \right] \, dx.
\]

We show in the sequel that the first term is a degenerate U-statistic and contributes with the variance in the limiting distribution, while the second gives the asymptotic bias. In turn, assumption A3 ensures that the third and fourth terms are negligible. To begin with, observe that the first moment of \( r_n(x, X) \) reads

\[
E_X [r_n(x, X)] = \varphi^{1/2}(x) \int_X K_{hn}(x - X) f(X) \, dX
\]

\[
= \varphi^{1/2}(x) \int_u K(u) f(x + uh_n) \, du
\]

\[
= \varphi^{1/2}(x) \int_u K(u) \left[ f(x) + \frac{1}{2} f'(x) uh_n + f''(x) u^2 h_n^2 \right] \, du
\]

\[
= \varphi^{1/2}(x) f(x) + O(h_n^2),
\]

where \( f^{(i)}(\cdot) \) denotes the \( i \)-th derivative of \( f(\cdot) \) and \( x^* \in [x, x + uh_n] \). Applying similar algebra to the second moment yields

\[
E_X [r_n^2(x, X)] = h_n^{-1} e_K \varphi(x) f(x) + O(1).
\]

This means that

\[
E(I_{2n}) = \frac{1}{n} \int_x E_X [r_n^2(x, X)] \, dx - \frac{1}{n} \int_x E_X [r_n(x, X)] \, dx
\]

\[
= \frac{1}{n} \int_x \left[ h_n^{-1} e_K \varphi(x) f(x) + O(1) \right] \, dx + O(n^{-1})
\]

\[
= n^{-1} h_n^{-1} e_K \int_x \varphi(x) f(x) \, dx + O(n^{-1}),
\]

whereas \( \text{Var}(I_{2n}) = O(n^{-3} h_n^{-2}) \). It then follows from Chebyshev’s inequality that

\[
nh_n^{1/2} I_{2n} - h_n^{-1/2} e_K E[\varphi(x)] = o_p(1).
\]

In turn, the deterministic term \( I_{3n} \) is proportional to the integrated squared bias of the fixed kernel density estimation, hence it is of order \( O(h_n^4) \). Assumption A3 then implies that

\[
n h_n^{1/2} I_{3n} = o(1).
\]

Further,

\[
E(I_{4n}) = 2 \int_x \varphi(x) E_X \left[ \hat{f}(x) - E\hat{f}(x) \right] \left[ E\hat{f}(x) - f(x) \right] \, dx = 0,
\]

25
whereas $E(I_{kn}^2) = O\left(n^{-1}h_n^4\right)$ as in Hall (1984, Lemma 1). It then suffices to impose assumption A3 to ensure that $nh_n^{1/2}I_{1n} = o_p(1)$ by Chebyshev’s inequality. Lastly, recall that $I_{1n} = \sum_{i<j} H_n(X_i, X_j)$, where

$$H_n(X_i, X_j) = 2n^{-2} \int \hat{r}_n(x, X_i)\hat{r}_n(x, X_j) \, dx.$$ 

As $H_n(X_i, X_j)$ is symmetric, centered and such that $E[H_n(X_i, X_j) | X_j] = 0$ almost surely, $I_{1n}$ is a degenerate U-statistic. Lemma 3 then establishes that $nh_n^{1/2}I_{1n} \xrightarrow{d} N(0, V_H)$, where

$$V_H = \frac{n^4 h_n}{2} E_{X_1, X_2} [H_n^2(X_1, X_2)]$$

$$= 2h_n \int_{X_1, X_2} [\int_x \hat{r}_n(x, X_1)\hat{r}_n(x, X_2) \, dx]^2 f(X_1, X_2) \, d(x_1, x_2)$$

$$= 2h_n \int_{x,y} \left[ \int_x \hat{r}_n(x, X)\hat{r}_n(y, X) \, f(X) \, dX \right]^2 \, d(x, y)$$

$$= 2 \int_x \varphi(x) \varphi(x + v h_n) \left[ \int_u K(u)K(u + v) f(x - uh_n) \, du \right]^2 \, d(x, v)$$

$$\simeq 2 \int_x \varphi^2(x) \left[ \int_u K(u)K(u + v) f(x - uh_n) \, du \right]^2 \, d(x, v)$$

which completes the proof.

**Proof of (14).** Consider the following von Mises expansion

$$\Phi_{f,h}(\gamma) = \Theta_{f+\gamma h} = \int_S \left[ f(x, \theta_x) - f(x) - \gamma h(x) \right]^2 \left[ f(x) + \gamma h(x) \right] \, dx,$$

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

$$\frac{\partial \Phi_{f,h}(\gamma)}{\partial \gamma} = 2 \int_S \left[ f(x, \theta_x) - f(x) - \gamma h(x) \right] \left[ f(x) + \gamma h(x) \right] \, dx$$

$$- 2 \int_S \left[ f(x, \theta_x) - f(x) - \gamma h(x) \right] \left[ f(x) + \gamma h(x) \right] h(x) \, dx$$

$$+ \int_S \left[ f(x, \theta_x) - f(x) - \gamma h(x) \right]^2 h(x) \, dx.$$

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 $D\Phi_f = \frac{\partial}{\partial \gamma} \Phi_{f,h}(0)$ is singular. In turn, the second functional derivative reads

$$\frac{\partial^2 \Phi_{f,h}(\gamma)}{\partial \gamma \partial \gamma'} = 2 \int_S \left[ f(x, \theta_x) - f(x) - \gamma h(x) \right] \left[ f(x, \theta_x) - f(x) - \gamma h(x) \right] h(x) \, dx$$
\[+ 2 \int_S \frac{\partial f(x, \theta_\gamma)}{\partial \theta} \frac{\partial^2 f(x, \theta_\gamma)}{\partial \gamma^2} [f(x, \theta_\gamma) - f_x - \gamma h_x] \, dx\]
\[+ 2 \int_S \frac{\partial f(x, \theta_\gamma)}{\partial \theta} \frac{\partial f(x, \theta_\gamma)}{\partial \theta} \frac{\partial^2 f(x, \theta_\gamma)}{\partial \gamma^2} [f_x + \gamma h_x] \, dx\]
\[- 4 \int_S \frac{\partial f(x, \theta_\gamma)}{\partial \theta} \frac{\partial f(x, \theta_\gamma)}{\partial \gamma} [f_x + \gamma h_x] \, dx\]
\[+ 4 \int_S \frac{\partial f(x, \theta_\gamma)}{\partial \theta} \frac{\partial f(x, \theta_\gamma)}{\partial \gamma} \frac{\partial f(x, \theta_\gamma)}{\partial \gamma} \frac{\partial f(x, \theta_\gamma)}{\partial \gamma} \, dx\]
\[+ 2 \int_S \frac{f_x + \gamma h_x}{\gamma} \, dx - 4 \int_S \frac{f(x, \theta_\gamma) - f_x - \gamma h_x}{\gamma} \, dx\]

which reduces to (14) by evaluating at \( \gamma = 0 \) and imposing the null.

**Proof of Proposition 2.** Let \( H(x) \equiv \hat{F}(x) - F(x) \), where \( \hat{F}(\cdot) \) is the pointwise kernel estimate of the true cumulative distribution function \( F(\cdot) \). Under the null, the following functional Taylor expansion is valid

\[\Phi_{\theta_{f+h}} = \int_{x,y} I(x \in S) \left[ t^D_f(x, y) + f(x) \delta(y - x) \right] dH(x) dH(y) + O(\|h_x\|^3),\]

where \( t^D_f \) is a continuous functional that includes the first and second terms of (14) as well as the regular part of its third term, and \( \delta(\cdot) \) is a Dirac mass at zero (Schwartz, 1966). By replacing \( h(x) \) by \( \hat{f}(x) - f(x) \), it is readily seen that

\[\int_{x,y} I(x \in S) t^D_f(x, y) dH(x) dH(y)\]

is negligible since it converges at rate \( n \) to a sum of independent \( \chi^2 \) distributions (Serfling, 1980). In turn, applying Lemma 1 with \( \varphi(x) = I(x \in S) f(x) \) yields that

\[\int_{x,y} I(x \in S) f(x) \delta(y - x) dH(x) dH(y) = \int_S f(x) h(x) \, dH(x)\]

converges in distribution at rate \( nh_n^{1/4} \) to a Gaussian variate with mean \( b_n^{-1/4} \delta_G \) and variance \( \sigma_G^2 \).

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

\[\hat{r}_n^D - \frac{nh_n^{1/2}}{\sigma_D} \sum_{i=1}^n I(x_{in} \in S) \frac{[f(x_{in}, \theta_f) - f(x_{in})]^2}{d(x_{in})} \xrightarrow{d} N(0,1),\]

27
Proof of (27). Computing the second differential of the expression above with which recovers (25) if evaluated at \( n \)

\[
\Phi_{f^{[n]}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(x_{in} \in S) \left[ f^{[n]}(x_{in}, \theta_{f^{[n]}}) - f^{[n]}(x_{in}) \right]^2 \\
= E \left\{ \mathbb{I}(x_{1n} \in S) \left[ f^{[n]}(x_{1n}, \theta_{f^{[n]}}) - f^{[n]}(x_{1n}) \right]^2 \right\} + o_p \left( n^{-1/2} \right) \\
= \epsilon_n^2 E \left\{ \mathbb{I}(x_{1n} \in S) \ell_D^2(x_{1n}) \right\} + o_p \left( n^{-1/2} h_n^{-1/2} \right) \\
= n^{-1} h_n^{-1/2} \epsilon_n^2 + o_p \left( n^{-1} h_n^{-1/2} \right).
\]

Observe that there is no asymptotic cost in estimating the parameter vector \( \theta \) provided that the sequence of local alternatives approaches the null hypothesis at a slower rate than \( \sqrt{n} \) as in (18). Applying a similar argument to the gamma kernel version completes the proof (see the proof of Proposition 7).

**Proof of (25).** Consider the following von Mises expansion

\[
\Lambda_{f,h}(\gamma) = \Lambda_{f+\gamma h} = \int_S \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right]^2 \left[ f(x) + \gamma h(x) \right] dx,
\]

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

\[
\frac{\partial \Lambda_{f,h}(\gamma)}{\partial \gamma} = 2 \int_S \frac{\partial \mathcal{H}_{\theta_\gamma}(x)}{\partial \theta} \frac{\partial \theta_\gamma}{\partial \gamma} \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right] \left[ f(x) + \gamma h(x) \right] dx \\
- 2 \int_S \frac{\partial \mathcal{H}_{f+\gamma h}(x)}{\partial \gamma} \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right] \left[ f(x) + \gamma h(x) \right] dx \\
+ \int_S \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right]^2 h(x) dx,
\]

which recovers (25) if evaluated at \( \gamma = 0 \).

**Proof of (27).** Computing the second differential of the expression above with respect to \( \gamma \) yields

\[
\frac{\partial^2 \Lambda_{f,h}(\gamma)}{\partial \gamma \partial \gamma'} = 2 \int_S \frac{\partial^2 \mathcal{H}_{\theta_\gamma}(x)}{\partial \theta \partial \theta'} \frac{\partial \theta_\gamma}{\partial \gamma} \frac{\partial \theta_\gamma}{\partial \gamma'} \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right] \left[ f(x) + \gamma h(x) \right] dx \\
+ 2 \int_S \frac{\partial \mathcal{H}_{\theta_\gamma}(x)}{\partial \theta} \frac{\partial^2 \theta_\gamma}{\partial \gamma \partial \gamma'} \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right] \left[ f(x) + \gamma h(x) \right] dx \\
+ 2 \int_S \frac{\partial \mathcal{H}_{f+\gamma h}(x)}{\partial \gamma} \frac{\partial \theta_\gamma}{\partial \gamma} \frac{\partial \theta_\gamma}{\partial \gamma'} \left[ f(x) + \gamma h(x) \right] dx \\
- 4 \int_S \frac{\partial \mathcal{H}_{\theta_\gamma}(x)}{\partial \theta} \frac{\partial \theta_\gamma}{\partial \gamma} \frac{\partial \mathcal{H}_{f+\gamma h}(x)}{\partial \gamma} \left[ f(x) + \gamma h(x) \right] dx \\
+ 4 \int_S \frac{\partial \mathcal{H}_{f+\gamma h}(x)}{\partial \gamma} \frac{\partial \theta_\gamma}{\partial \gamma} \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right] h(x) dx \\
- 2 \int_S \frac{\partial^2 \mathcal{H}_{f+\gamma h}(x)}{\partial \gamma \partial \gamma'} \left[ \mathcal{H}_{\theta_\gamma}(x) - \mathcal{H}_{f+\gamma h}(x) \right] \left[ f(x) + \gamma h(x) \right] dx.
\]
\[
+ 2 \int_S \frac{\partial H_{f + \gamma h}}{\partial \gamma} \frac{\partial H_{f + \gamma h}}{\partial \gamma'} [f(x) + \gamma h(x)] \, dx \\
- 4 \int_S \frac{\partial H_{f + \gamma h}(x)}{\partial \gamma} [H_{\theta_0}(x) - H_{f + \gamma h}(x)] \, h(x) \, dx,
\]
which equals (27) for \( \gamma = 0 \).

**Proof of Proposition 5.** Under the null, the following functional Taylor expansion is valid
\[
\Lambda_{f + h} = \int_{x,y} \mathbb{I}(x \in S) \left[ \ell^H_f(x,y) + \delta(y - x) \mathcal{H}_f(x)/S_x \right] \, dH(x) \, dH(y) + O \left( \|h\|^3 \right),
\]
where \( \ell^H_f \) is a continuous functional encompassing the second and third terms of (27) as well as the regular part of its first term and \( S_x \) denotes the survival function \( 1 - F(x) \). Replacing \( h(x) \) by \( \hat{f}(x) - f(x) \) yields that the first term
\[
\int_{x,y} \mathbb{I}(x \in S) \ell^H_f(x,y) \, dH(x) \, dH(y)
\]
converges at a rate \( n \) and it is therefore negligible. In turn, applying Lemma 4 with \( \varphi(x) = \mathbb{I}(x \in S) \mathcal{H}_f(x)/S_x \) yields that
\[
\int_{x,y} \mathbb{I}(x \in S) \delta(y - x) \mathcal{H}_f(x)/S_x \, dH(x) \, dH(y) = \int_S \mathcal{H}_f(x)/S_x \left[ \hat{f}(x) - f(x) \right]^2 \, dx
\]
converges weakly at rate \( n b_n^{1/4} \) to a normal distribution with mean \( b_n^{-1/4} \lambda_H \) and variance \( \varsigma^2_H \).

**Proof of Proposition 6.** Consider the above functional Taylor expansion with \( h(x) = \hat{f}(x) - f(x) \). Once more, the first term converges at a rate \( n \), whereas Lemma 1 implies that
\[
\int_{x,y} \mathbb{I}(x \in S) \mathcal{H}_f(x)/S(x) \delta(y - x) \, dH(x) \, dH(y) = \int_S \mathcal{H}_f(x)/S(x) \left[ \hat{f}(x) - f(x) \right]^2 \, dx
\]
converges in distribution at rate \( n b_n^{1/4} \) to a normal variate with mean \( b_n^{-1/4} \lambda_G \) and variance \( \varsigma^2_G \).

**Proof of Proposition 7.** Again, 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 \( H_\lambda^n \) and assumptions A1 to A4,
\[
\tilde{\tau}_n^H - \frac{nb_n^{1/4}}{\varsigma_G} \sum_{i=1}^n \mathbb{I}(x_{in} \in S) \left| \mathcal{H}(x_{in}) - \mathcal{H}(x_{in}) \right|^2 \overset{d}{\longrightarrow} N(0,1).
\]
The result then follows from the fact that \( \xi_G \overset{p}{\to} \xi_G \) and
\[
\Lambda_{f[n]} = \frac{1}{n} \sum_{i=1}^{n} I(x_i \in S) \left[ H^{[n]}(x_i, \theta_{f[n]}) - H_f^{[n]}(x_i) \right]^2
\]
\[
= E \left\{ I(x_1 \in S) \left[ H^{[n]}(x_1, \theta_{f[n]}) - H_f^{[n]}(x_1) \right]^2 \right\} + O_p \left( n^{-1/2} \right)
\]
\[
= \varepsilon_n^2 E \left[ I(x_1 \in S) \ell_H(x_1) \right] + o_p \left( n^{-1} \right)
\]
\[
= n^{-1} b_n^{-1/4} \ell_n^H + o_p \left( n^{-1} \right).
\]
We omit the proof for the fixed kernel version of the H-test as it is completely analogous (see the proof of Proposition 3).

**Proof of Proposition 8.** The implicit functional corresponding to the M-estimator associated with the H-test is
\[
\int_S \frac{\partial \mathcal{H}(x, \theta_H)}{\partial \theta} \left[ \mathcal{H}(x, \theta_H) - \mathcal{H}_f(x) \right] f(x) \, dx = 0,
\]
which results in the following expansion
\[
\int_S \frac{\partial \mathcal{H}(x, \theta_H)}{\partial \theta} \left[ \mathcal{H}(x, \theta_H) - \mathcal{H}_{f + \gamma h}(x) \right] [f(x) + \gamma h(x)] \, dx = 0.
\]
Differentiating with respect to \( \gamma \) then entails
\[
\int_S \frac{\partial \mathcal{H}(x, \theta_H)}{\partial \theta} \left[ \mathcal{H}(x, \theta_H) - \mathcal{H}_{f + \gamma h}(x) \right] [f(x) + \gamma h(x)] \, dx = 0
\]
which recovers (32) if one evaluates at \( \gamma = 0 \). As the first term of the right-hand side of (26) converges at a slower rate than the second, (33) will drive the asymptotic distribution of \( \theta_H \). A straightforward application of Lemma 2 then yields \( n^{1/2}(\hat{\theta}_n^H - \theta_*) \overset{d}{\to} N(0, \Omega_H) \), where \( \theta_* \) is the pseudo-true value of the parameter vector that minimizes the distance between the nonparametric and parametric hazard estimates. Under the null hypothesis, \( \theta_* \) coincides with the true parameter vector \( \theta_0 \) and therefore \( \hat{\theta}_n^H \) is a consistent estimator. The same results applies under any sequence of local alternatives that approaches the null hypothesis at a slower rate than root-\( n \) as in (30).

**Proof of Proposition 9.** We demonstrate the result only for the H-test since the proof for the D-test follows in similar fashion. We must show that, under the null,
\[
\Lambda_f(\hat{\phi}) = \frac{1}{n} \sum_{i=1}^{n} I(\hat{e}_i \in S) \left[ \mathcal{H}_\phi(\hat{e}_i) - \mathcal{H}_f(\hat{e}_i) \right]^2
\]
has the same limiting distribution as its counterpart $\Lambda_f(\phi_0)$ in (24). We start by pursuing a third-order Taylor expansion with Lagrange remainder

$$
\Lambda_f(\hat{\phi}) = \Lambda_f(\phi_0) + \Lambda'_f(\phi_0)(\hat{\phi} - \phi_0) + \frac{1}{2}\Lambda''_f(\phi_0)(\hat{\phi} - \phi_0, \hat{\phi} - \phi_0)
+ \Lambda'''_f(\phi_*) (\hat{\phi} - \phi_0, \hat{\phi} - \phi_0, \hat{\phi} - \phi_0)
= \Lambda_f(\phi_0) + \Delta_1 + \Delta_2 + \Delta_3,
$$

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

$$
\Lambda'_f(\phi_0) = 2\int_S [\mathcal{H}_f(e) - \mathcal{H}_e(e)] [\mathcal{H}_f'(e) - \mathcal{H}_e'(e)] f(e) \, de
+ \int_S [\mathcal{H}_f(e) - \mathcal{H}_e(e)]^2 f'(e) \, de,
$$

where all differentials are with respect to $\phi$ evaluated at $\phi_0$. Under the null hypothesis, $\Lambda'_f(\phi_0) = 0$ and $\Lambda'_f(\phi_0) = O_p(n^{-1}h_n^{-1})$ given that $(\hat{f} - f)^2 = O_p(n^{-1}h_n^{-1})$ and $(\hat{f}' - f')^2 = O_p(n^{-1}h_n^{-3})$. Thus, the first term $\Delta_1$ is of order $O_p(n^{-3/2}h_n^{-1})$. Similarly, $\Lambda''_f(\phi_0) = O_p(n^{-1}h_n^{-3})$ and $\Delta_2 = O_p(n^{-2}h_n^{-3})$. The last term requires more caution for it is not evaluated at the true parameter $\phi_0$. However, it is not difficult to show that

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

so that $\Delta_3 = O_p\left(n^{-2}h_n^{-7/2}\right) + O_p\left(n^{-5/2}h_n^{-3}\right)$. Under the assumption A3, it follows that $nh_n^{1/2}(\Delta_1 + \Delta_2 + \Delta_3) = o_p(1)$, since $nh_n^{1/2} \Delta_1 = o_p(n^{-2/5})$, $nh_n^{1/2} \Delta_2 = o_p(n^{-1/2})$, and $nh_n^{1/2} \Delta_3 = o_p(n^{-2/5}) + o_p(n^{-1})$. This implies that the limiting distributions of $\Lambda_f(\hat{\phi})$ and $\Lambda_f(\phi_0)$ coincide and hence the H-test is nuisance parameter free. For the gamma kernel version of the H-test, the same argument applies if one replaces $h_n$ with the bandwidth $b_n$ for gamma kernel estimation.

31
References


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


Bouezmarni, T., Scaillet, O., 2002, Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, Université Catholique de Louvain and Université de Genève.


Fernandes, M., Grammig, J., 2003, A family of asymmetric autoregressive conditional duration models, Graduate School of Economics, Getulio Vargas Foundation.

Fernandes, M., Monteiro, P. K., 2003, Central limit theorem for asymmetric kernel functionals, Graduate School of Economics, Getulio Vargas Foundation.


Grammig, J., Maurer, K.-O., 2000, Non-monotonic hazard functions and the autoregressive conditional duration model, Econometrics Journal 3, 16–38.


Figure 1: Finite-sample properties of the tests
Data generating process: Exponential ACD

Figure 2: Finite-sample properties of the tests
Data generating process: Weibull ACD
Figure 3: Finite-sample properties of the tests
Data generating process: Burr ACD
Figure 4: Finite-sample properties of the tests
Data generating process: Generalized gamma ACD
Figure 5: Instantaneous volatility implied by the ACD models for a typical Exxon trading day (September, 9, 1999)
Table 1
Descriptive statistics of Exxon price durations

<table>
<thead>
<tr>
<th>duration</th>
<th>sample size</th>
<th>mean</th>
<th>overdispersion</th>
<th>Q(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>plain</td>
<td>2,717</td>
<td>513.0</td>
<td>1.405</td>
<td>131.6</td>
</tr>
<tr>
<td>seasonally adjusted</td>
<td>2,717</td>
<td>1.000</td>
<td>1.196</td>
<td>68.2</td>
</tr>
</tbody>
</table>

Plain durations refer to the time interval needed to observe a cumulative change in the mid-price of at least $0.125, whereas seasonally adjusted durations consider a cubic spline estimate of the intraday seasonal component. Overdispersion stands for the ratio between standard deviation and mean. Q(10) denotes the Ljung-Box statistic of order 10.

Table 2
Test results for the different linear ACD models

<table>
<thead>
<tr>
<th>distribution</th>
<th>testing procedure</th>
<th>p-value in sample</th>
<th>p-value out of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>exponential</td>
<td>overdispersion test</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.09)</td>
</tr>
<tr>
<td></td>
<td>D-test with optimal kernel</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>D-test with gamma kernel</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>H-test with gamma kernel</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>D-test for log-residuals</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Weibull</td>
<td>overdispersion test</td>
<td>0.00 (0.02)</td>
<td>0.00 (0.10)</td>
</tr>
<tr>
<td></td>
<td>D-test with optimal kernel</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>D-test with gamma kernel</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>H-test with gamma kernel</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>D-test for log residuals</td>
<td>0.00 (0.00)</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>Burr</td>
<td>D-test with optimal kernel</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>D-test with gamma kernel</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>H-test with gamma kernel</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>D-test for log residuals</td>
<td>0.12 (0.14)</td>
<td>0.05 (0.08)</td>
</tr>
<tr>
<td>generalized gamma</td>
<td>D-test with optimal kernel</td>
<td>0.48</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>D-test with gamma kernel</td>
<td>0.47</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>H-test with gamma kernel</td>
<td>0.65</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>D-test for log-residuals</td>
<td>0.56 (0.83)</td>
<td>0.26 (0.20)</td>
</tr>
</tbody>
</table>

Figures in parentheses correspond to asymptotic p-values, whereas the others correspond to p-values based on the empirical distribution of the test statistic stemming from 499 artificial bootstrap samples.


466. INFLAÇÃO E FLEXIBILIDADE SALARIAL - Marcelo Côrtes Neri; Maurício Pinheiro – Dezembro de 2002 – 16 págs.


469. EMPLOYMENT AND PRODUCTIVITY IN BRAZIL IN THE NINETIES - José Márcio Camargo; Marcelo Côrtes Neri; Mauricio Cortez Reis – Dezembro de 2002 – 32 págs.

<table>
<thead>
<tr>
<th>No.</th>
<th>Título</th>
<th>Autor/a(s)</th>
<th>Data</th>
<th>Páginas</th>
</tr>
</thead>
<tbody>
<tr>
<td>471</td>
<td>CUSTO DE CICLO ECONÔMICO NO BRASIL EM UM MODELO COM RESTRIÇÃO A CRÉDITO</td>
<td>Bárbara Vasconcelos Boavista da Cunha; Pedro Cavalcanti Ferreira</td>
<td>Janeiro de 2003</td>
<td>21 págs</td>
</tr>
<tr>
<td>472</td>
<td>THE COSTS OF EDUCATION, LONGEVITY AND THE POVERTY OF NATIONS</td>
<td>Pedro Cavalcanti Ferreira; Samuel de Abreu Pessoa</td>
<td>Janeiro de 2003</td>
<td>31 págs</td>
</tr>
<tr>
<td>473</td>
<td>A GENERALIZATION OF JUDD’S METHOD OF OUT-STEADY-STATE COMPARISONS IN PERFECT FORESIGHT MODELS</td>
<td>Paulo Barelli; Samuel de Abreu Pessoa</td>
<td>Fevereiro de 2003</td>
<td>7 págs</td>
</tr>
<tr>
<td>474</td>
<td>AS LEIS DA FALÊNCIA: UMA ABORDAGEM ECONÔMICA</td>
<td>Aloísio Pessoa de Araújo</td>
<td>Fevereiro de 2003</td>
<td>25 págs</td>
</tr>
<tr>
<td>475</td>
<td>THE LONG-RUN ECONOMIC IMPACT OF AIDS</td>
<td>Pedro Cavalcanti G. Ferreira; Samuel de Abreu Pessoa</td>
<td>Fevereiro de 2003</td>
<td>30 págs</td>
</tr>
<tr>
<td>476</td>
<td>A MONETARY MECHANISM FOR SHARING CAPITAL: DIAMOND AND DYBVIG MEET KIYOTAKI AND WRIGHT</td>
<td>Ricardo de O. Cavalcanti</td>
<td>Fevereiro de 2003</td>
<td>16 págs</td>
</tr>
<tr>
<td>477</td>
<td>INADA CONDITIONS IMPLY THAT PRODUCTION FUNCTION MUST BE ASYMPTOTICALLY COBB-DOUGLAS</td>
<td>Paulo Barelli; Samuel de Abreu Pessoa</td>
<td>Março de 2003</td>
<td>4 págs</td>
</tr>
<tr>
<td>478</td>
<td>TEMPORAL AGGREGATION AND BANDWIDTH SELECTION IN ESTIMATING LONG MEMORY</td>
<td>Leonardo R. Souza</td>
<td>Março de 2003</td>
<td>19 págs</td>
</tr>
<tr>
<td>479</td>
<td>A NOTE ON COLE AND STOCKMAN</td>
<td>Paulo Barelli; Samuel de Abreu Pessoa</td>
<td>Abril de 2003</td>
<td>8 págs</td>
</tr>
<tr>
<td>480</td>
<td>A HIPÓTESE DAS EXPECTATIVAS NA ESTRUTURA A TERMO DE JUROS NO BRASIL: UMA APLICAÇÃO DE MODELOS DE VALOR PRESENTE</td>
<td>Alexandre Maia Correia Lima; João Victor Issler</td>
<td>Maio de 2003</td>
<td>30 págs</td>
</tr>
<tr>
<td>481</td>
<td>ON THE WELFARE COSTS OF BUSINESS CYCLES IN THE 20TH CENTURY</td>
<td>João Victor Issler; Afonso Arinos de Mello Franco; Osmani Teixeira de Carvalho Guillén</td>
<td>Maio de 2003</td>
<td>29 págs</td>
</tr>
<tr>
<td>482</td>
<td>RETORNOS ANORMAIS E ESTRATÉGIAS CONTRÁRIAS</td>
<td>Marco Antonio Bonomo; Ivana Dall’Agnol</td>
<td>Junho de 2003</td>
<td>27 págs</td>
</tr>
<tr>
<td>483</td>
<td>EVOLUÇÃO DA PRODUTIVIDADE TOTAL DOS FATORES NA ECONOMIA BRASILEIRA: UMA ANÁLISE COMPARATIVA</td>
<td>Victor Gomes; Samuel de Abreu Pessoa; Fernando A. Veloso</td>
<td>Junho de 2003</td>
<td>45 págs</td>
</tr>
<tr>
<td>484</td>
<td>MIGRAÇÃO, SELEÇÃO E DIFERENÇAS REGIONAIS DE Renda no Brasil</td>
<td>Enestor da Rosa dos Santos Junior; Naércio Menezes Filho; Pedro Cavalcanti Ferreira</td>
<td>Junho de 2003</td>
<td>23 págs</td>
</tr>
<tr>
<td>485</td>
<td>THE RISK PREMIUM ON BRAZILIAN GOVERNMENT DEBT, 1996-2002</td>
<td>André Soares Loureiro; Fernando de Holanda Barbosa</td>
<td>Junho de 2003</td>
<td>16 págs</td>
</tr>
<tr>
<td>486</td>
<td>FORECASTING ELECTRICITY DEMAND USING GENERALIZED LONG MEMORY</td>
<td>Lacir Jorge Soares; Leonardo Rocha Souza</td>
<td>Junho de 2003</td>
<td>22 págs</td>
</tr>
</tbody>
</table>


493. REAL EXCHANGE RATE MISALIGNMENTS - Maria Cristina T. Terra; Frederico Estrella Carneiro Valladares – Agosto de 2003 – 26 págs.

494. ELASTICITY OF SUBSTITUTION BETWEEN CAPITAL AND LABOR: A PANEL DATA APPROACH - Samuel de Abreu Pessoa ; Silvia Matos Pessoa; Rafael Rob – Agosto de 2003 – 30 págs.


