ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Nonparametric Kernel Testing in Semiparametric Autoregressive Conditional Duration Model

Pipat Wongsaart and Jiti Gao

September 2011

Working Paper 18/11

Nonparametric Kernel Testing in Semiparametric Autoregressive Conditional Duration Model

Pipat Wongsaart and Jiti Gao[1]

Monash University, Australia

A crucially important advantage of the semiparametric regression approach to the nonlinear autoregressive conditional duration (ACD) model developed in Wongsaart et al. (2011), i.e. the so–called Semiparametric ACD (SEMI–ACD) model, is the fact that its estimation method does not require a parametric assumption on the conditional distribution of the standardized duration process and, therefore, the shape of the baseline hazard function. The research in this paper complements that of Wongsaart et al. (2011) by introducing a nonparametric procedure to test the parametric density function of ACD error through the use of the SEMI–ACD based residual. The hypothetical structure of the test is useful, not only to the establishment of a better parametric ACD model, but also to the specification testing of a number of financial market microstructure hypotheses, especially those related to the information asymmetry in finance. The testing procedure introduced in this paper differs in many ways from those discussed in existing literatures, for example Aït-Sahalia (1996), Gao and King (2004) and Fernandes and Grammig (2005). We show theoretically and experimentally the statistical validity of our testing procedure, while demonstrating its usefulness and practicality using datasets from New York and Australia Stock Exchange.

JEL Classification: C14, C41, F31.

keyword Duration model, hazard rates and random measures, nonparametric kernel testing.

1. Introduction

A well known property of the so–called high–frequency data in finance is the fact that market events are clustered over time. This suggests that financial durations, i.e. the inter–event waiting times, may follow positively an autocorrelated process with strong persistence. This feature may be captured in a number of alternative ways through different econometric methods based on duration, intensity or counting representations of a point process. Today, one of the most well–known approaches in the literature is the ACD model introduced by Engle and Russell (1998); see also Engle and Russell (1997) for an application of the model to foreign exchange data. The ACD model considers a stochastic process that is simply a sequence of times $\{i_0, i_1, \ldots, i_n, \ldots\}$ with $i_0 < i_1 < \cdots < i_n \ldots$. The interval between

[1]Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East VIC 3145, Australia. Email: jiti.gao@monash.edu

two arrival times, i.e. $x_t = i_t - i_{t-1}$, measures the length of times commonly known as the durations by which $\{x_t\}$ is a nonnegative stationary process adapted to the filtration $\{\mathcal{F}_t, t \in \mathbb{Z}\}$ with $\mathcal{F}_t$ representing the previous history. The ACD class of models assumes a multiplicative model of x_t of the form

$$x_t = \psi_t \varepsilon_t, \tag{1.1}$$

where $\{\varepsilon_t\}$ is an independent and identically distributed (i.i.d.) innovation series with density $f(\varepsilon; \phi)$ (with non-negative support and $E[\varepsilon_1] = 1$), $\psi_t \equiv \vartheta(x_{t-1}, \ldots, x_{t-p}, \psi_{t-1}, \ldots, \psi_{t-q})$ and $\vartheta : \mathbb{R}_+^p \times \mathbb{R}_+^q \to \mathbb{R}_+$ is a strictly positive–valued function.

Expression (1.1) suggests that there is now a host of potential specifications for the ACD model where each is defined by different specifications for the expected durations and for the distribution of ε. While a number of existing studies examine some generalizations and hypothesis testing of the former, for example Fernandes and Grammig (2006) and Meitz and Teräsvirta (2006), the misspecification of the baseline distribution may have quite serious implications. When a data generating process is based a non–monotonic baseline hazard rate function, Grammig and Maurer (2000) show that the quasi maximum likelihood estimation fails to provide sound finite sample results even in quite large sample cases. Furthermore, 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 (Prigent et al. (2000)). Moreover, Drost and Werker (2004) argue against the i.i.d. assumption in (1.1) in favor of a semiparametric alternative that allows the distribution function of the innovations to be dependent on the past. The resulting model relies heavily on the linear parameterization of the conditional duration and the assumption that it is correctly specified.

Therefore, one of the benefits of the SEMI–ACD model developed in Wongsaart et al. (2011) is the fact that imposition of such distribution assumption is not required in the model's estimation procedure. Furthermore, a three–step modeling procedure, which was also suggested in conjunction with the SEMI–ACD model, enables a straightforward method of empirically estimating the the density (and therefore the survival and baseline hazard functions) of the innovations. Nonetheless, for such results to be advantageous to empirical and theoretical studies of financial market microstructure, there should be a method of gauging their closeness to those of existing distributions in the literature.

This paper presents a two–step semiparametric procedure to test the marginal density function of durations. While the objective of the first step is to model the dynamics of the financial duration process using the above mentioned SEMI–ACD model, the second step tests the SEMI–ACD residual about the baseline density of the standardized duration.

Regarding the hypothesis testing in the second step, Aït-Sahalia (1996) introduces a nonparametric testing procedure to test the marginal density functions of a class of diffusion processes under the β-mixing condition. Fernandes and Grammig (2005) extend Aït–Sahalia's approach to hypothesis testing in the context of a parametric ACD model. Unlike Fernandes and Grammig (2005), whose work focuses only on addressing the boundary bias that arises as durations have a support which is bounded from below, the research of this paper concentrates also on the importance of bandwidth selection in nonparametric kernel testing, while an extra measure is taken to minimize the impact of the bias induced by the kernel estimation. Both Aït-Sahalia (1996) and Fernandes and Grammig (2005) select the bandwidths of their tests by simply using an adjusted version of the Silverman (1986) rule of thumb. However, existing studies, e.g. Gao and King (2004) and Gao (2007), show that in fact bandwidth selection in nonparametric kernel testing is not a straightforward matter.

Generally speaking, one can distinguish in the literature two approaches to deal with this bandwidth parameter choice in nonparametric and semiparametric kernel methods used for constructing model specification tests. The first approach is to use an estimation–based optimal bandwidth value, such as a cross–validation bandwidth. However, this may lead to a poor performance of the test in finite sample studies because the estimation–based optimal bandwidth may not necessarily imply that the corresponding test is optimal. The second approach is to consider among a set of pre–specified suitable values for the bandwidth.

In this paper, we extend a method that is first initiated in Horowitz and Spokoiny (2001) for testing of a parametric model of a conditional mean function against a nonparametric alternative. The idea of the test is to consider simultaneously a family of test statistics associated with H_T, which represents a set of bandwidth values. The proposed test rejects the null hypothesis if at least one of the test statistics for $h \in H_T$ is sufficiently large. For the reasons that will be explained in Section 3, Horowitz and Spokoiny (2001) define this test as an *adaptive and rate optimal test*. More recently, Gao and King (2004) extend Horowithz and Spokoiny's approach to the parametric specification testing of the marginal density in a continuous–time diffusion model.

The most important feature about the two–step SEMI–ACD estimation and nonparametric specification testing introduced in this paper that clearly differentiates it from the work of Gao and King (2004) is the fact that each of these steps is implemented based on estimates, i.e. use of the algorithm based conditional durations in the SEMI–ACD estimation and testing the SEMI–ACD based residuals about the parametric marginal density function of the standardized duration. Therefore, the main contribution of this paper is to construct such testing procedure and to show theoretically and empirically its asymptotic costlessness.

Furthermore, the theoretical illustration of such asymptotic costlessness differs significantly from what appears in Fernandes and Grammig (2005). Using the fact that the first step of their two–step procedure is to estimate the conditional duration process by the quasi maximum likelihood (QML) estimation, in their theoretical discussion, Fernandes and Grammig simply assume a root-n consistency of their parametric estimator of the standardized duration. The use of the SEMI–ACD model suggests that, we first establish the consistency of the algorithm–based estimate of the standardized duration. Then an important implication of such consistency is shown on the unaffected limit distribution of the test statistic. We will elaborate further on the differences between the research in this paper and that of Fernandes and Grammig (2005) in Section 3 below.

Additionally, let us note the potential usefulness of our testing procedure that resides in its applicability to various first–step ACD estimations by which the consistency of their estimates of the standardized duration can be established. The most obvious example is in Figure 2 of Drost and Werker (2004), which compares the estimated density and the standard exponential density of their ACD innovations. In this case our testing procedure can be used to test the statistical suitability of the standard exponential density for Drost and Werker's ACD innovation.

The remainder of this section summarizes a number of notable findings and key contributions of the research in this paper.

- This paper deals with a new ACD model for the case where $\{\psi_t\}$ is semiparametric and $\{\varepsilon_t\}$ is a stationary time series and its distribution is nonparametrically unknown. The testing procedure developed in this paper displays a strong consistency against a sequence of local alternatives, i.e. reasonable power values which are gradually incremental toward one, even with relatively small distances between the null and alternative hypotheses as well as sample sizes. Neither Fernandes and Grammig (2005) nor Gao and King (2004) presents the empirical evidence about the consistency of their test statistics against a sequence of local alternatives.

- Both the experimental and empirical studies show that the newly introduced testing procedure is statistically powerful in the sense that it is able to gauge the mixture of distributions even for cases by which the distributions belong to the same family. As discussed in detail in Section 5, the test statistic suggests that a mixture weibull–gamma distribution is able to best describe the price duration processes in question.

- Through the use of the SEMI–ACD residuals, evidently the procedure introduced in this paper is able to successfully overcome the latency problem, which arises because of the unobservability of the standardized duration in practice.

The remainder of this paper is organized as follows. Section 2 discusses the statistical consistency of the algorithm–based estimation of the standardized durations. Section 3 discusses the testing of the marginal density and a number of new asymptotic results. Section 4 provides experimental evidence to demonstrate the statistical validity and usefulness of our testing procedure. Section 5 applies the procedure to test parametric density functions for price duration at the New York Stock Exchange (NYSE) and the Australian Stock Exchange (ASX). Finally, Section 6 summarizes the main results and offers concluding remarks. For ease of exposition, Appendices A and B collect all underlying assumptions. All proofs and technical lemmas are given in Appendix C of the supplementary document.

2. Algorithm–Based Estimation of the Standardized Durations

An important feature of our two–step semiparametric estimation and testing procedure is the fact that each of these steps is implemented based on an estimate, which is computed using the iterative estimation algorithm studied in detail in Wongsaart et al. (2011). The purpose of this section is to discuss statistical consistency of the algorithm–based estimation of the standardized duration, which constitutes the first step of our procedure, in order to pave way for the theoretical development of our test statistic in Section 3.

Let us begin our discussion in this section with a brief review of the so–called SEMI–ACD model. The main focus of this paper is on the SEMI–ACD(1,1) model of the form

$$\psi_t \equiv \gamma x_{t-1} + g\left(\psi_{t-1}\right), \tag{2.1}$$

where $g(\cdot)$ satisfies Assumption B.1 below. The strict stationarity and ergodicity assumed in Engle and Russell (1998) implies that Assumption B.1 holds for the ACD(1,1) model. To derive the estimators of γ and g, observe that the multiplicative model in (1.1) can be written in terms of an additive noise of the form

$$x_t = \gamma x_{t-1} + g\left(\psi_{t-1}\right) + \eta_t \tag{2.2}$$

where $\eta_t = \psi_t\left(\varepsilon_t - 1\right)$ and $\{\varepsilon_t\}$ is a sequence of positive stationary errors with $E\left[\varepsilon_1\right] = 1$ and $E\left[\varepsilon_1^{4+\delta}\right] < \infty$ for some $\delta > 0$. We then have under the assumption that $E[\eta_t|\psi_{t-1}] = 0$

$$g\left(\psi_{t-1}\right) = E\left[x_t|\psi_{t-1}\right] - \gamma E\left[x_{t-1}|\psi_{t-1}\right] = g_1\left(\psi_{t-1}\right) - \gamma g_2\left(\psi_{t-1}\right). \tag{2.3}$$

The fact that the conditional duration process is not observable in practice suggests that the usual one–step partially linear autoregressive estimation cannot be applied in the SEMI–ACD case. To address this latency problem the estimation procedure employed in Wongsaart et al. (2011) is based on an algorithmically computed estimate of the t-th conditional duration at the m^*-th iteration defined by

$$\widehat{\psi}_{t,m^*} \equiv \widehat{\gamma}_{m^*}(h)x_{t-1} + \widehat{g}_{1,h}(\widehat{\psi}_{t-1,m^*-1}) - \widehat{\gamma}_{m^*}(h)\widehat{g}_{2,h}(\widehat{\psi}_{t-1,m^*-1}), \tag{2.4}$$

where $\widehat{\gamma}_{m^*}(h)$ is the kernel weighted LS estimate of γ at the m^*-th iteration, m^* is a pre-specified maximum number of iterations and

$$\widehat{g}_{j,h}(\widehat{\psi}_{t-1,m^*-1}) = \sum_{s=m^*+\iota}^{T} W_{s,h}(\widehat{\psi}_{t-1,m^*-1})x_{s-j+1} \tag{2.5}$$

for $j = 1, 2$ and $\iota \in \mathbb{N}$, where $W_{s,h}$ is a probability weight function of the form

$$W_{s,h}(\widehat{\psi}_{t-1,m^*-1}) = \frac{k_h(\widehat{\psi}_{t-1,m^*-1} - \widehat{\psi}_{s-1,m^*-1})}{\sum_{s=m^*+\iota}^{T} k_h(\widehat{\psi}_{t-1,m^*-1} - \widehat{\psi}_{s-1,m^*-1})}$$

with $k_h\left(\cdot\right) = h^{-1}k\left(\cdot/h\right)$, k is a real-valued kernel function satisfying Assumption B.3 in Appendix B and $h = h_T \in H_T$.

In the SEMI–ACD case, the kernel based least squares estimators of γ and $\sigma^2 = E(\eta_1^2)$ are written as

$$\widehat{\gamma}_{\widehat{\psi}}(h) = \left\{ \sum_{t=1}^{T} \widehat{u}_{t+1}^2 \omega(\widehat{\psi}_{t,m^*}) \right\}^{-1} \left\{ \sum_{t=1}^{T} \widehat{u}_{t+1} \left(x_{t+1} - \widehat{g}_{1,h}(\widehat{\psi}_{t,m^*}) \right) \omega(\widehat{\psi}_{t,m^*}) \right\} \tag{2.6}$$

and

$$\widehat{\sigma}^2_{\widehat{\psi}}(h) = \frac{1}{T} \sum_{t=1}^{T} \{ x_{t+1} - \widehat{\gamma}_{\widehat{\psi}}(h) x_t - \widehat{g}_{1,h}(\widehat{\psi}_{t,m^*}) + \widehat{\gamma}_{\widehat{\psi}}(h) \widehat{g}_{2,h}(\widehat{\psi}_{t,m^*}) \}^2 \omega(\widehat{\psi}_{t,m^*}), \tag{2.7}$$

where $\widehat{u}_{t+1} = x_t - \widehat{g}_{2,h}(\widehat{\psi}_{t,m^*})$, $\widehat{\overline{g}}_h = g(\psi_t) - \widehat{g}_h(\widehat{\psi}_{t,m^*})$ and $\omega(\cdot)$ is a known nonnegative weight function satisfying Assumption B.3 in Appendix B.

In order to proceed with the hypothesis test in the second step, we first must introduce an algorithm–based estimate of the standardized duration, which is the SEMI–ACD residual

$$\widehat{\varepsilon}_{t,m^*} = \frac{x_t}{\widehat{\psi}_{t,m^*}}, \tag{2.8}$$

where $\widehat{\psi}_{t,m^*}$ is the algorithm–based estimate of the conditional duration as defined in (2.4). In the study of the nuisance parameter freeness of their test statistic, Fernandes and Grammig (2005) use the fact that their first step estimation is based on the Engle and Russell's parametric ACD model, hence simply assume the so–called root-N consistency of their estimate for the unobserved ε. The use of the SEMI–ACD model suggests that we must follow quite a different route in this paper.

To study theoretically the impact of the SEMI–ACD's algorithm–based estimation, here we rely on the transformation

$$|\widehat{\varepsilon}_{t,m^*} - \varepsilon_t| = \left| \frac{x_t}{\widehat{\psi}_{t,m^*}} - \frac{x_t}{\psi_t} \right| = \left\{ \frac{\varepsilon_t}{\psi_t} \right\} \left| \widehat{\psi}_{t,m^*} - \psi_t \right| \left\{ \frac{\psi_t}{\widehat{\psi}_{t,m^*}} \right\}. \tag{2.9}$$

Hence, a uniform consistency of $\widehat{\psi}_{t,m^*}$, for example, should immediately lead to a similar mode of consistency of $\widehat{\varepsilon}_{t,m^*}$. Both the establishment and the statistical consistency of the iterative estimation algorithm are discussed in detail in Wongsaart et al. (2011), therefore are not the main concern of this paper. However, in the discussion that follows, in order to establish the consistency of our test statistic against various alternatives, we establish the uniform consistency of $\widehat{\psi}_{t,m^*}$ and $\widehat{\varepsilon}_{t,m^*}$ as a technical lemma, which is presented in Appendix C of the supplementary document.

3. Testing Marginal Density Function

This section discusses the basic construction and a number of important asymptotic properties of the test statistic for testing the marginal densities. For the sake of convenience in introducing the new testing procedure, its underlying motivations and statistical results, and also in order to highlight the fact that the test may be used for testing the density of a stationary random variable in general, Section 3.1 proceeds under the assumption that standardized durations are observable. This assumption will be relaxed in Section 3.2.

3.1. Testing Procedure with Observable Durations

Let $\{\varepsilon_t\}$ be the standardized duration process of a financial event and let $f(\cdot)$ and $f(\cdot, \theta)$ be a nonparametric and a parametric forms of the marginal density function of $\{\varepsilon_t\}$, respectively. Furthermore, let Θ denote a parameter space in R^q and $\theta_0 \in \Theta$ denote the true value of θ. We consider in this paper a testing procedure for testing the null hypothesis

$$\mathcal{H}_0 : f(\varepsilon) = f(\varepsilon, \theta_0) \tag{3.1}$$

against a sequence of local alternatives

$$\mathcal{H}_1 : f(\varepsilon) = f(\varepsilon, \theta_1) + C_T \Delta_T(\varepsilon), \quad \theta_1 \in \Theta, \tag{3.2}$$

where $0 \leq C_T \leq 1$, $\lim_{T \to \infty} C_T = 0$ and $\Delta_T(\varepsilon)$ is a continuous function that is chosen to satisfy $\int \Delta_T(\varepsilon) d\varepsilon = 0$. In this case, $\Delta_T(\varepsilon)$ must be constructed in the way that the alternative function is still a probability density under $\mathcal{H}_1$. The analysis in this paper considers the case where

$$\Delta_T(\varepsilon) = \varphi(\varepsilon) - f(\varepsilon, \theta_1) \tag{3.3}$$

so that the alternative hypothesis in (3.2) can be rewritten as a *semiparametric* mixture density

$$\mathcal{H}_1 : f(\varepsilon) = (1 - C_T) f(\varepsilon, \theta_1) + C_T \varphi(\varepsilon), \tag{3.4}$$

where φ denotes a nonparametric density. Clearly, a couple of special cases of such a structure of a sequence of local alternatives are the global alternatives of the forms

$$\mathcal{H}_1' : f(\varepsilon) = f(\varepsilon, \theta_1) \quad \text{and} \quad \mathcal{H}_1'' : f(\varepsilon) = (1 - C) f(\varepsilon, \theta_1) + C \varphi(\varepsilon) \tag{3.5}$$

that are obtained for cases by which $C_T = 0$ and $C_T = C$ for $0 < C < 1$, respectively.

The usefulness of such specification testing in an ACD class of models is best evidenced when considering the conditional intensity function of arrival times, which is commonly defined in the literature as

$$\lambda(i|N(i), i_1, \ldots, i_{N(i)}) = \lambda_0 \left(\frac{i - i_{N(i)}}{\psi_{N(i)+1}} \right) \frac{1}{\psi_{N(i)+1}}, \tag{3.6}$$

where $\lambda_0(\varepsilon) = \frac{f(\varepsilon;\phi)}{S(\varepsilon;\phi)}$ is the baseline hazard and $S(\varepsilon;\phi) = \int_\varepsilon^\infty f(u;\phi)du$ is the survivor function. Clearly, the shape of this type of accelerated failure time model depends very much on that of the baseline hazard $\lambda_0(\cdot)$. For example, if it is assumed that the durations are conditionally exponential so that the baseline hazard is simply one, and the conditional intensity is then

$$\lambda(i|N(i), i_1, \ldots, i_{N(i)}) = \psi_{N(i)+1}^{-1}. \tag{3.7}$$

Hence, an idiosyncrasy of the Engle and Russell's EACD and WACD models is that the implied hazard functions conditional on past durations are restricted to be either constant, increasing or decreasing with respect to duration. A number of studies, such as Zhang et al. (2001) and Bauwens and Veredas (2004), raise questions about the appropriateness of imposing such restrictions. Furthermore, Grammig and Maurer (2000) present an experimental evidence that suggests the misspecification of the hazard function can severely deteriorate the model's ability to predict expected durations.

A crucially important advantage of the semiparametric modeling procedure developed in Wongsaart et al. (2011) is the fact that a specific distribution assumption of the innovation ε, and therefore the above mentioned restrictions, is not required for their so–called SEMI–ACD model to be consistently and efficiently estimated. Therefore, the specification testing procedure presented in this paper complements the estimation method proposed in Wongsaart et al. (2011) in the sense that the distribution of the standardized duration can now be statistically determined by the SEMI–ACD based estimates. On the one hand, such knowledge is very useful in empirical and theoretical studies of financial market microstructure, while on the other hand it can be used as guideline to building a better parametric ACD model; see Section 5 below for an empirical illustration and further discussion.

Moreover, such hypothesis testing is particularly useful when taking into account the existence of the information asymmetry in financial markets whereby various types of traders, for example informed and uninformed traders, may co–exist. The differences in their trading behavior lead us to believe that a mixture distribution might be a useful model for waiting–time distribution in finance; see also De Luca and Gallo (2004). Therefore, not only the hypothesis testing against such an alternative as (3.2) enables an empirical test of the information asymmetry hypothesis, it also provides additional information on whether the use of mixture distribution in a parametric ACD model is essential.

The main idea behind the test statistic considered in this paper is to compare a consistent nonparametric density estimator directly to a parametric density in question. To discuss our test statistic, let us first define the distance function

$$D(f,\theta) = \int (f(\varepsilon) - f(\varepsilon,\theta))^2 f(\varepsilon) d\varepsilon. \tag{3.8}$$

There are at least two alternative methods of estimating $D(f,\theta)$ considered in the literature. The first alternative is based on the estimator

$$D(\widehat{f},\tilde{\theta}) = \int (\widehat{f}(\varepsilon) - f(\varepsilon,\tilde{\theta}))^2 \widehat{f}(\varepsilon) d\varepsilon, \tag{3.9}$$

where $\widehat{f}(\varepsilon) = (1/T)\sum_{t=1}^{T} k_h(\varepsilon - \varepsilon_t)$ is the standard kernel density estimator, $k_h(\cdot) = h^{-1}k(\cdot/h)$, $k(\cdot)$ is a kernel function and $\tilde{\theta}$ is a consistent estimator of θ, while the second alternative is to use

$$\tilde{D}(\widehat{f},\tilde{\theta}) = \int (\widehat{f}(\varepsilon) - \tilde{f}(\varepsilon,\tilde{\theta}))^2 \widehat{f}(\varepsilon) d\varepsilon, \tag{3.10}$$

where

$$\tilde{f}(\varepsilon,\tilde{\theta}) = \sum_{t=1}^{T} w_t(\varepsilon) f(\varepsilon_t,\tilde{\theta}) \tag{3.11}$$

is a nonparametric estimator of $f(\varepsilon,\theta)$,

$$w_t(\varepsilon) = w_t(\varepsilon,h) = (1/T)k_h(\varepsilon - \varepsilon_t) \times \left[\frac{(s_2(\varepsilon) - s_1(\varepsilon)(\varepsilon - \varepsilon_t))}{(s_2(\varepsilon)s_0(\varepsilon) - s_1^2(\varepsilon))}\right]$$

and $s_r(\varepsilon) = (1/T)\sum_{s=1}^{T} k_h(\varepsilon - \varepsilon_s)(\varepsilon - \varepsilon_s)^r$ for $r = 0,1,2$. Note that, in (3.11), $f(\varepsilon_t,\tilde{\theta})$ is properly smoothed in order to cancel out the bias involved in $\widehat{f}(\varepsilon)$.

As shown in Gao and King (2004), a suitably standardized version of $\tilde{D}(\widehat{f},\tilde{\theta})$ is better both theoretically and empirically than that of $D(\widehat{f},\tilde{\theta})$.

Thus, the test statistic that will be the main focus of this paper is based on $\tilde{D}(\widehat{f},\tilde{\theta})$ in (3.10) and is written as

$$\widehat{N}_T = \widehat{N}_T(h) = Th \int (\hat{f}(\varepsilon) - \tilde{f}(\varepsilon,\tilde{\theta}))^2 \hat{f}(\varepsilon) d\varepsilon. \tag{3.12}$$

In this case, Gao and King (2004) show that under $\mathcal{H}_0$,

$$L_T(h) = \frac{\widehat{N}_T(h) - \mu_0}{\sqrt{h}\sigma_0} \to_D N(0,1) \;\; \text{as} \;\; T \to \infty, \tag{3.13}$$

where $\mu_0 = R(k) \int_{\infty}^{\infty} f^2(\varepsilon) d\varepsilon$ and $\sigma_0^2 = 2k^{(4)}(0) \int f^4(\varepsilon) d\varepsilon$, and that

$$\widehat{L}_T(h) = \frac{\widehat{N}_T(h) - \widehat{\mu}_T(h)}{\sqrt{h}\widehat{\sigma}_T(h)} \to_D N(0,1) \;\; \text{as} \;\; T \to \infty, \tag{3.14}$$

where $\widehat{\mu}_T(h) = R(k) \cdot \left(\frac{1}{T}\sum_{i=1}^T \widehat{f}(\varepsilon_t)\right)$ and $\widehat{\sigma}_T^2(h) = 2k^{(4)}(0) \cdot \left(\frac{1}{T}\sum_{t=1}^T \widehat{f}^3(\varepsilon_t)\right)$, in which $R(k) = \int k^2(u)du < \infty$ and $k^{(j)}(0)$ denotes the j-times convolution product of $k(\cdot)$ given by $k^{(4)}(0) = \int_{-\infty}^{\infty} L^2(\varepsilon)d\varepsilon$ with $L(\varepsilon) = \int_{-\infty}^{\infty} k(y)k(\varepsilon + y)dy$.

Furthermore, suppose that there is a random data–driven $\widehat{h}$ such that $(\widehat{h}/h) - 1 \to_P 0$ as $T \to \infty$. Then, we have under $\mathcal{H}_0$

$$\widehat{L}_T(\widehat{h}) = \frac{\widehat{N}_T(\widehat{h}) - \widehat{\mu}_T(\widehat{h})}{\sqrt{\widehat{h}}\widehat{\sigma}_T(\widehat{h})} \to_D N(0,1) \;\; \text{as} \;\; T \to \infty. \tag{3.15}$$

In this case, the asymptotic normality as stated in (3.13) to (3.15) can be obtained under a conventional condition of the form $\limsup_{T\to\infty} Th^5 < \infty$.

For the implementation of their proposed test within the context of the parametric ACD specifications, Fernandes and Grammig (2005) suggest that an undersmoothing–adjusted theoretically optimal bandwidth can be used. Although one may argue in favor the use of such a rule of thumb or other estimation–based optimal bandwidth this may contribute to a poor performance of the test in finite sample studies because the estimation–based bandwidth may not necessarily imply that the corresponding test is optimal.

To address this problem, the current paper suggests using

$$L^* = \max_{h \in H_T} \widehat{L}_T(h) = \max_{h \in H_T} \frac{\widehat{N}_T(h) - \widehat{\mu}_T(h)}{\sqrt{h}\widehat{\sigma}_T(h)}, \tag{3.16}$$

where it is assumed that H_T is finite with J_T number of elements. A specific formulation of H_T employed here is a geometric grid of the form

$$H_T = \{h = h_{\max} a^k : h \geq h_{\min}, \; k = 0, 1, 2, \ldots\}, \tag{3.17}$$

where $0 < h_{\min} < h_{\max}$ and $0 < a < 1$. In this case, $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$. More detailed conditions on $h_{\max}$ and $h_{\min}$ are given in Appendix A.

3.2. Testing Procedure Based on the SEMI–ACD Estimates

For notational clarity, hereafter let $L_{T,\varepsilon}(h) \equiv L_T(h)$, $\widehat{L}_{T,\varepsilon}(h) \equiv \widehat{L}_T(h)$ and $L_\varepsilon^* \equiv L^*$. The substitution of the algorithm–based estimate $\widehat{\varepsilon}_{t,m^*}$ for ε_t, suggests that $\widehat{L}_{T,\varepsilon}(h)$ can be reformulated as

$$\widehat{L}_{T,\widehat{\varepsilon}}(h) = \frac{\widehat{N}_{T,\widehat{\varepsilon}}(h) - \widehat{\mu}_{T,\widehat{\varepsilon}}(h)}{\sqrt{h}\widehat{\sigma}_{T,\widehat{\varepsilon}}(h)}, \tag{3.18}$$

so that the test statistic L_ε^* in (3.16) is now

$$L_{\widehat{\varepsilon}}^* = \max_{h \in H_T} \widehat{L}_{T,\widehat{\varepsilon}}(h) = \max_{h \in H_T} \frac{\widehat{N}_{T,\widehat{\varepsilon}}(h) - \widehat{\mu}_{T,\widehat{\varepsilon}}(h)}{\sqrt{h}\widehat{\sigma}_{T,\widehat{\varepsilon}}(h)}, \tag{3.19}$$

where

$$\widehat{N}_{T,\widehat{\varepsilon}}(h) = \frac{Th}{T} \sum_{n=1}^{T} (\widehat{f}(\widehat{\varepsilon}_{t,m^*} - \tilde{f}(\widehat{\varepsilon}_{t,m^*}, \tilde{\theta}))^2 = h \sum_{t=1}^{T} (\widehat{f}(\widehat{\varepsilon}_{t,m^*}) - \tilde{f}(\widehat{\varepsilon}_{t,m^*}, \tilde{\theta}))^2, \tag{3.20}$$

in which $\tilde{\theta}$ is the estimate of the true value θ_0, based on the residual $\widehat{\varepsilon}_{t,m^*}$.

An essential result to ensure the asymptotic triviality of such use of SEMI–ACD estimate, and therefore the consistency of the test introduced in this paper, is to establish that

$$\widehat{L}_{T,\widehat{\varepsilon}}(h) = \widehat{L}_{T,\varepsilon}(h) + R_T(h), \tag{3.21}$$

where $R_T(h)$ is the remainder that converges to zero in probability as $T \to \infty$. The proof of (3.21) relies heavily on the uniform convergence of $\widehat{\psi}_{t,m^*}$ and $\widehat{\varepsilon}_{t,m^*}$, which is discussed in Section 2. In view of (3.21), Lemma B.6 of Gao and King (2004) already shows that

$$\widehat{L}_{T,\varepsilon}(h) = L_{T,\varepsilon}(h) + o_P\{1\}. \tag{3.22}$$

Therefore, the asymptotic results of $L_T(h)$ and $\max_{h \in H_T} \widehat{L}_{T,\varepsilon}(h)$ established in Gao and King (2004) remain valid for $\widehat{L}_{T,\widehat{\varepsilon}}(h)$ and $L_{\widehat{\varepsilon}}^*$. Appendix A lists the underlying assumptions, which are required in order to establish some relevant asymptotic results in Gao and King (2004).

Let us now introduce the method of computing a critical value for $L_{\widehat{\varepsilon}}^*$. For $0 < \alpha < 1$, the exact α-level critical value, l_α, is defined as the $1 - \alpha$ quantile of the exact finite sample distribution of $L_{\widehat{\varepsilon}}^*$. However, since θ_0 is unknown, l_α cannot be evaluated in practice. Therefore, to implement our testing procedure, we suggest choosing a simulated α-level critical value, l_α^*, by using the following simulation scheme, which can be employed to both re–samples of the sampled data or generated data from a known marginal density.

Step 3.1: Perform SEMI–ACD estimation to obtain $\widehat{\varepsilon}_{t,m^*}$.

Step 3.2: The true value θ_0 is estimated based on $\widehat{\varepsilon}_{t,m^*}$. The resulting estimate is then denoted by $\widehat{\theta}$.

Step 3.3: Compute $L^*_{\widehat{\varepsilon}}$ based on $\widehat{\varepsilon}_{t,m^*}$ and $\widehat{\theta}$.

Step 3.4: Repeat the preceding steps Q times in order to obtain Q versions of $L^*_{\widehat{\varepsilon}}$, i.e. $L^*_{\widehat{\varepsilon},q}$ for $q = 1, 2, \ldots, Q$. The simulated critical value l^*_α is then the $(1-\alpha)$ percentile of the Q values of $L^*_{\widehat{\varepsilon}}$.

The following result is essential in order to ensure the statistical validity of the above simulation scheme; see also the discussion in Remark 3.1 regarding the mathematical proof of these results.

Theorem 3.1. *Let Assumptions A.1–A.7 listed in Appendix A and B.1–B.4 listed in Appendix B hold. Then, under $\mathcal{H}_0$ we have*

$$\lim_{n\to\infty} P(L^*_{\widehat{\varepsilon}} > l^*_\alpha) = \alpha. \tag{3.23}$$

Furthermore, to ensure the statistical validity of using the residual based test statistic $L^*_{\widehat{\varepsilon}}$ to test the parametric marginal density function of the unobserved standardized duration, ε, we must also establish its consistency against the fixed and a sequence of local alternatives introduced previously.

3.2.1. Consistency of the Test against a Fixed Alternative

The purpose of this section is to show that $L^*_{\widehat{\varepsilon}}$ is consistent against such fixed alternatives as in (3.5). Let $\mathcal{F} = \{f(\cdot,\theta) : \theta \in \Theta\}$ be a set of density functions that satisfy Assumption A.3 in Appendix A and $F(\theta) = (f(\varepsilon_1,\theta),\ldots,f(\varepsilon_T,\theta))^\tau$ and $\bar{f} = (f(\varepsilon_1),\ldots,f(\varepsilon_T))^\tau$. Then the distance between f and $\mathcal{F}$ can be measured by the following normalized l_2 distance $\rho(f,\mathcal{F}) = \left[\inf_{\theta\in\Theta}\left(\frac{1}{T}\|\bar{f} - F(\theta)\|^2\right)\right]^{1/2}$, where $\|\cdot\|$ denotes the Euclidean norm.

Theorem 3.2. *Let Assumptions A.1–A.7 listed in Appendix A and B.1–B.4 listed in Appendix B hold. In addition, if there is a $C_\rho > 0$ such that $\lim_{T\to\infty} P(\rho(f,\mathcal{F}) \geq C_\rho) = 1$ holds, then*

$$\lim_{T\to\infty} P(L^*_{\widehat{\varepsilon}} > l^*_\alpha) = 1. \tag{3.24}$$

Theorem 3.2 shows that if $\mathcal{H}_0$ is false, then $\rho(f, \mathcal{F}) \geq \mathcal{C}_\rho$ for all sufficiently large T and some $C_\rho > 0$. A consistent test will reject a false $\mathcal{H}_0$ with probability approaching one as $T \to \infty$.

3.2.2. Consistency of the Test against a Sequence of Local Alternatives

The purpose of this section is to show that $L^*_{\widehat{\varepsilon}}$ is consistent against a sequence of of local alternatives of the form

$$f_T(\varepsilon) = f(\varepsilon, \theta_1) + C_T \Delta_T(\varepsilon) \tag{3.25}$$

whose distance from the parametric model converges to zero at the rate determined by $C_T \geq C_0 T^{-1/2}\sqrt{\log\log T}$ for some $C_0 > 0$ and $\theta_1 \in \Theta$. For convenience, let

$$\bar{\Delta}_T = (\Delta_T(\varepsilon_1), \ldots, \Delta_T(\varepsilon_T))^\tau \quad \text{and} \quad \bar{f}_T = (f_T(\varepsilon_1), \ldots, f_T(\varepsilon_T))^\tau.$$

We can now write the l_2 distance

$$\frac{1}{T}\|\bar{f}_T - F(\theta_1)\|^2 = \frac{C_T^2}{T}\|\bar{\Delta}_T\|^2 = C_T^2 \left(\frac{1}{T}\sum_{t=1}^{T} |\Delta_T(\varepsilon_t)|^2\right). \tag{3.26}$$

To ensure that the rate of convergence of $\tilde{f}_T$ to the parametric model $F(\theta_1)$ is the same as the rate of convergence of C_T to zero, we assume that for some $\delta > 0$ $\Delta_T(\varepsilon)$ is a continuous function which is normalized so that

$$\lim_{T\to\infty} P\left(\frac{1}{T}\sum_{t=1}^{T} |\Delta_T(\varepsilon_t)|^2 \geq \delta\right) = 1. \tag{3.27}$$

Theorem 3.3. *Let Assumptions A.1–A.7 listed in Appendix A and B.1–B.4 listed in Appendix B hold. Let f_T satisfy (3.25) with $C_T \geq C T^{-1/2}\sqrt{\log\log T}$ for some constant $C > 0$. In addition, let condition (3.27) hold. Then*

$$\lim_{T\to\infty} P(L^*_{\widehat{\varepsilon}} > l^*_\alpha) = 1. \tag{3.28}$$

Theorem 3.3 implies a similar conclusion to what has been discussed in Theorem 3.2.

Remark 3.1. An important step to completing the mathematical proof of Theorems 3.1 to 3.3 is to establish the asymptotic result presented in (3.21). In view of (3.21), the results of the theorems can then be conveniently obtained using Lemmas B.6 and B.7 of Gao and King (2004). While the technical assumptions required in the establishment of these results are presented in Appendix B below for convenience, a more detailed discussion of the proof can be found in Appendix C of the supplementary document.

Sections 4 and 5 below examine the finite–sample performance of the proposed estimation and testing procedure.

4. Finite Sample Properties: Monte Carlo Studies

The objective of the Monte Carlo exercises conducted in this section is threefold. Firstly, it is to provide some experimental evidence on the asymptotic optimality of the Cross Validation (CV) bandwidth selection in the estimation of the SEMI–ACD model. Although the theoretical discussion of this result is presented in detail in Wongsaart et al. (2011), this is the first time that an experimental evidence in support of such a procedure is provided. Secondly, it is to assess the performance of our hypothesis test in finite samples. Thirdly, it is to compare the finite sample performance of our testing procedure to that of other methods in order to shed some further light on the problem of bandwidth selection in nonparametric kernel testing. These methods are discussed in further detail below.

All computations in this section are done in Splus. While the specific details about each exercise are given in a relevant sub–section below, let us introduce here the so–called Mackey–Glass ACD (MG–ACD) model, which is employed in this section as a model example.

To facilitate the experimental studies in this section, we generate a MG–ACD model by which the dynamics of the duration process is described as

$$\psi_t = \gamma x_{t-1} + \lambda \left(\frac{\psi_{t-1}}{1 + \psi_{t-1}^2} \right) \tag{4.1}$$

with $\gamma = 0.5$ and $\lambda = 0.75$, respectively. The MG–ACD model is suitable for our exercises for a number of reasons. The most important reason being the fact that the above given functional form of g ensures that the simulated duration process $\{x_t\}$ is strictly stationary; see Section 2.4 of Tjøstheim (1994) for detail. Secondly, it can be shown using Lemma 3.4.4 and Theorem 3.4.10 of Györfi et al. (1989) that the resulting conditional duration process $\{\psi_t\}$ is α-mixing. Hence, the dynamics of the simulated duration process is completely explained by the conditional duration, which is essentially the key assumption behind most of the ACD class of models.

4.1. Asymptotic Optimality of Bandwidth Selection in the SEMI–ACD Model

To construct an adaptive data–driven estimation without unnecessary complication, let us begin our discussion with a usual case by which the conditional duration is assumed to be observable. For $1 \leq n \leq N = T-1$, the leave-one-out estimators of g_j can be defined as

$$\widehat{g}_{j,n}(\psi_n) = \frac{1}{N-1}\sum_{s\neq n}\frac{K_h(\psi_n-\psi_s)x_{x+2-j}}{\widehat{f}_{h,n}(\psi_n)}, \tag{4.2}$$

and let $\widehat{g}_{h,n}(\psi_n) = \widehat{g}_{1,n}(\psi_n) - \gamma\widehat{g}_{2,n}(\psi_n)$, where $\widehat{f}_{h,n}(\psi_n) = \frac{1}{N-1}\sum_{s\neq n}K_h(\psi_n-\psi_s)$. The leave-one-out estimate $\widetilde{\gamma}_\psi(h)$ of γ is defined by minimizing $\sum_{n=1}^{N}\{x_{n+1}-\gamma x_n - \widehat{g}_{h,n}(\psi_n)\}^2\omega(\psi_n)$. Hence, the CV function in this case can be written as

$$CV_\psi(h) = \frac{1}{N}\sum_{n=1}^{N}\{x_{n+1}-\widetilde{\gamma}_\psi(h)x_n - \widehat{g}_{1,n}(\psi_n) + \widetilde{\gamma}_\psi(h)\widehat{g}_{2,n}(\psi_n)\}^2\omega(\psi_n), \tag{4.3}$$

by which an optimal value $\widehat{h}_{C,\psi}$ of h is chosen such that

$$CV(\widehat{h}_{C,\psi}) = \inf_{h\in H_N} CV(h). \tag{4.4}$$

Unobservability of the conditional duration in practice suggests that we replace ψ_n with the algorithm–based estimate $\widehat{\psi}_{n,m^*}$. This leads to the ASE and the CV functions:

$$D_{\widehat{\psi}}(h) = \frac{1}{N}\sum_{n=1}^{N}\left\{[\widetilde{\gamma}_{\widehat{\psi}}(h)x_n + \widehat{g}_h^*(\widehat{\psi}_{n,m^*})] - [\gamma x_n + g(\psi_n)]\right\}^2 \omega(\widehat{\psi}_{n,m^*}), \tag{4.5}$$

where $\widehat{g}_h^*(\widehat{\psi}_{n,m^*}) = \widehat{g}_{1,h}(\widehat{\psi}_{n,m^*}) - \widetilde{\gamma}_{\widehat{\psi}}(h)\widehat{g}_{2,h}(\widehat{\psi}_{n,m^*})$, and

$$CV_{\widehat{\psi}}(h) = \frac{1}{N}\sum_{n=1}^{N}\{x_{n+1}-\widetilde{\gamma}_{\widehat{\psi}}(h)x_n - \widehat{g}_{1,n}(\widehat{\psi}_{n,m^*}) + \widetilde{\gamma}_{\widehat{\psi}}(h)\widehat{g}_{2,n}(\widehat{\psi}_{n,m^*})\}^2\,\omega(\widehat{\psi}_{n,m^*}). \tag{4.6}$$

In the context of the SEMI–ACD model, the cross–validation criterion consists of selecting the value $\widehat{h}_{C,\widehat{\psi}}$ of h that achieves

$$CV_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}}) = \inf_{h\in H_N} CV_{\widehat{\psi}}(h). \tag{4.7}$$

Assuming that some suitable assumptions hold, Wongsaart et al. (2011) show that the adaptive nonparametric prediction algorithm is asymptotically optimal in the sense that

$$\frac{D_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})}{D_\psi(h_0)} \longrightarrow_P 1, \quad \text{where} \quad h_0 = \arg\min_{h\in H_N} D_\psi(h). \tag{4.8}$$

To establish an experimental evidence to illustrate that the bandwidth selected using the rule in (4.7) is asymptotically optimal in the light of (4.8) we only have to demonstrate that

$$\frac{D_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})}{D_\psi(\widehat{h}_{C,\psi})} \longrightarrow_P 1 \quad \text{as} \quad T\to\infty. \tag{4.9}$$

The fact that the true data generating process is known in our simulation study suggests that we are not only able to compute $D_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})$ and $CV_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})$, but also $D_{\psi}(\widehat{h}_{C,\psi})$ and $CV_{\psi}(\widehat{h}_{C,\psi})$.

Furthermore, in order to establish an empirical evidence to illustrate the consistency of $\widehat{\varepsilon}_{t,m^*}$, we compute the average squared error

$$D_{\widehat{\varepsilon}}(h) = \frac{1}{T}\sum_{t=1}^{T}\{\widehat{\varepsilon}_{t,m^*} - \varepsilon_t\}^2 \tag{4.10}$$

for all number of observations considered, namely $T = 250, 750, \quad \text{and} \quad 1,400$.

Mathematically, the innovation ε may follow any distributional function such that the probability of it being less than zero is zero. However, to be consistent with the experimental exercises that follow, we assume in this sub–section that $\{\varepsilon_t\} \sim \text{Gamma}(\alpha, \theta)$. A detailed discussion on the gamma and other distributions considered in this paper is available below.

Table 4.1. Asymptotic optimality of bandwidth selection in SEMI–ACD model.

T	$D_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})$	$D_{\psi}(\widehat{h}_{C,\psi})$	r_D	$CV_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})$	$CV_{\psi}(\widehat{h}_{C,\psi})$	r_{CV}	$D_{\widehat{\varepsilon}}(h)$
250	0.0073	0.0065	1.1415	0.3664	0.3666	1.0003	0.0087
750	0.0047	0.0056	1.0485	0.3771	0.3788	0.9968	0.0063
1,400	0.0039	0.0052	1.0008	0.3745	0.3758	0.9963	0.0053

Note: In the above table, let $r_D = \frac{D_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})}{D_{\psi}(\widehat{h}_{C,\psi})}$ and $r_{CV} = \frac{CV_{\widehat{\psi}}(\widehat{h}_{C,\widehat{\psi}})}{CV_{\psi}(\widehat{h}_{C,\psi})}$.

Table 4.1 presents the simulation results about the asymptotic optimality of bandwidth selection in SEMI–ACD estimation. The forth column of Table 4.1 presents the results of r_D for each number of observations being considered. Clearly, the asymptotic optimality condition in (4.9) is well supported by these results. In this case, the resulting r_D have tendency to converge to 1 as T increases. $r_D = 1.0008$ at $T = 1,400$ is very close to 1. Furthermore, $D_{\widehat{\varepsilon}}(h)$, which is reported in the last column of the table, has also shown a strong tendency to converge to zero as $T \to \infty$. This is a convincing experimental evidence in support of our theory about the consistency of $\widehat{\varepsilon}_{t,m^*}$ in Section 2.

4.2. Nonparametric Specification Tests

In this section, we perform finite sample studies on the size and power of the test studied above against a number of alternatives as outlined in Section 3. To examine the robustness of the test, we attempt to formulate models with data generating processes that exhibit hazard rate functions of different shapes. To achieve this objective, here we employ a family of three–parameter generalized gamma densities given by

$$f_{GG}(\varepsilon;\alpha,\theta,\delta)=\left\{\frac{\delta\varepsilon^{\delta\alpha-1}}{\theta^{\delta\alpha}\Gamma(\alpha)}\right\}\exp\left\{-\left(\frac{\varepsilon}{\theta}\right)^{\delta}\right\} \quad \text{for} \quad \varepsilon\geq 0, \tag{4.11}$$

where $\alpha>0$, $\theta>0$ and $\delta>0$. These may include, among others, the gamma distribution ($\delta=1$), the exponential distribution ($\alpha=1$ and $\delta=1$) and the weibull distribution ($\alpha=1$), whose probability density functions can be, respectively, written as

$$f_{G}(\varepsilon;\theta,\alpha)=\left\{\frac{\varepsilon^{\alpha-1}}{\theta^{\alpha}\Gamma(\alpha)}\right\}\exp\left\{-\left(\frac{\varepsilon}{\theta}\right)\right\}, \quad f_{E}(\varepsilon;\lambda)=\lambda\exp\{-\lambda\varepsilon\} \tag{4.12}$$

and

$$f_{W}(\varepsilon;\theta,\delta)=\frac{\delta}{\theta}\left\{\frac{\varepsilon}{\theta}\right\}^{\delta-1}\exp\left\{-\left(\frac{\varepsilon}{\theta}\right)^{\delta}\right\},$$

where $\lambda=\frac{1}{\theta}$.

Glaser (1980) derives a well–known result which examines the sign of $\eta'=\frac{-f'(\varepsilon)}{f(\varepsilon)}$ in order to determine the shape of its hazard rate function. To apply Glaser's result, let us first define the following functions:

$$\eta'_{GG}(\varepsilon)=\frac{-f'_{GG}(\varepsilon)}{f_{GG}(\varepsilon)}=\varepsilon^{-2}\left[(\delta\alpha-1)+\delta(\delta-1)\left(\frac{\varepsilon}{\theta}\right)^{\delta}\right], \tag{4.13}$$

$$\eta'_{G}(\varepsilon)=\frac{-f'_{G}(\varepsilon)}{f_{G}(\varepsilon)}=\varepsilon^{-2}(\alpha-1), \tag{4.14}$$

$$\eta'_{E}(\varepsilon)=\frac{-f'_{E}(\varepsilon)}{f_{E}(\varepsilon)}=0. \tag{4.15}$$

A number of conclusions can then be drawn about the shape of the hazard rate functions of the generalized gamma, gamma and exponential distributions. Let us present some cases, which are useful for our studies, in Table 4.2.

Table 4.2. Shapes of the generalized gamma hazard rate functions.

Cases	Parameters	Hazard Rate Function
(1) $\delta\alpha-1<0$	$\delta=1$ and $\alpha<1$	Decreasing (Gamma Density)
	$\delta<1$ and $\alpha=1$	Decreasing (Weibull Density)
	$\delta>1$	U–Shaped
(2) $\delta\alpha-1>0$	$\delta=1$ and $\alpha>1$	Increasing (Gamma Density)
	$\delta>1$ and $\alpha=1$	Increasing (Weibull Density)
	$\delta<1$	Inverted U–Shaped
(3) $\delta\alpha-1=0$	$\delta=1$ and $\alpha=1$	Constant (Exponential Density)

To provide experimental evidence on the effectiveness of the test based on $\{\widehat{\varepsilon}_{t,m^*}\}$ instead of $\{\varepsilon_t\}$, we provide the results for each version of the test in two tables. The first table presents rejection rates for the marginal density test when implemented on a random sample generated under either the null (size) or the alternative (power) hypothesis, i.e. the test statistic being considered in this case is L^*_{ε}. These results are compared to those of the second table that contains the rejection rates when the tests are implemented on the SEMI–ACD residuals. That is, we use the random sample generated for the first table to simulate the MG–ACD process, estimate the SEMI–ACD model to obtain $\widehat{\varepsilon}_{t,m^*}$ and then apply our testing procedure. Therefore, the test statistic being considered in this case is $L^*_{\widehat{\varepsilon}}$.

Throughout the simulation, to compute the nonparametric estimators involved, we choose the normal kernel function given by

$$k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{4.16}$$

Furthermore, H_T is as defined in (3.17) with $h_{\min} = T^{-\frac{11}{36}}$, $h_{\max} = 2(\log\log T)^{-1}$, and $a = \frac{35}{36}$. Three different sizes of sample, namely $T = 176, 301,$ and 501, are considered. The corresponding simulated critical values at the α-level are found by using the simulation scheme proposed in Section 3.2, which is implemented at $M = 1,000$. The number of simulations used in producing the proceeding tables is 200.

4.2.1. Size of the Test

In this section, we consider the following null hypotheses:

(a) $\mathcal{H}_0 : f(\varepsilon) = f_{GG}(\varepsilon, \theta_0)$ with $\theta_0 = (\alpha_0 = 2, \delta_0 = 0.9, \beta_0 = 0.5)$, which suggests that the generalized gamma hazard rate function is *inverted U–shaped.*

(b) $\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0)$ with $\theta_0 = (\alpha_0 = 2, \beta_0 = 1)$, which suggests that the gamma hazard rate function is *monotonically increasing.*

(c) $\mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \lambda_0)$ with the rate parameter of $\lambda_0 = 0.5$.

Table 4.3. The size of the test L^*_{ε}.

	$\mathcal{H}_0 : f(\varepsilon) = f_{GG}(\varepsilon, \theta_0)$		$\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0)$		$\mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \lambda_0)$	
T	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
176	0.060	0.020	0.065	0.020	0.035	0.010
301	0.055	0.010	0.045	0.005	0.050	0.010
501	0.050	0.010	0.050	0.010	0.045	0.020

Table 4.4. The size of the test $L^*_{\hat{\varepsilon}}$.

T	$\mathcal{H}_0 : f(\varepsilon) = f_{GG}(\varepsilon, \theta_0)$		$\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0)$		$\mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \lambda_0)$	
	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
176	0.065	0.02	0.055	0.020	0.040	0.005
301	0.055	0.02	0.045	0.005	0.055	0.010
501	0.045	0.01	0.045	0.01	0.060	0.020

Tables 4.3 and 4.4 present simulation results on the size values of L^*_ε and $L^*_{\hat{\varepsilon}}$, respectively. In all cases, the rejection rates obtained are quite close to their corresponding critical levels. Furthermore, generally all the rates in Table 4.4 are similar to those of their ε based counterparts in Table 4.3. In Sections 4.2.2 and 4.2.3 below, we also introduce a measure, which enables this particular issue to be investigated in more detail.

4.2.2. Power of the Test against a Fixed Alternative

We examine in this section the power of the test in the following hypotheses

$$\mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \theta_0) \quad vs. \quad \mathcal{H}_a : f(\varepsilon) = f_G(\varepsilon, \theta_j) \tag{4.17}$$

$$\mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \theta_0) \quad vs. \quad \mathcal{H}_a : f(\varepsilon) = (1 - C_s) f_E(\varepsilon, \theta_1) + C_s \ \varphi(\varepsilon) \tag{4.18}$$

for $j = 1, 2, 3$ and $s = 1, 2, 3$, where $0 < C_s < 1$, θ_j are vectors of gamma parameters defined in the table below and $\varphi(\varepsilon)$ represents a nonparametric density function. The first set is constructed so that comparatively restrictive densities, with respect to the implied shape of the hazard rate function under $\mathcal{H}_0$, are tested against a set of more flexible densities. Specifically, we look at testing the null hypothesis of $f_E(\varepsilon, \theta_0)$ with a constant hazard rate function against alternatives $f_G(\varepsilon, \theta_j)$ for $j = 1, 2, 3$, whose hazard rate functions are incrementally increasing. With regard to the second set, to ensure that the alternative function in this case is still a probability density and that it is fairly close to the exponential law under $\mathcal{H}_0$, we let

$$\varphi(\varepsilon) = \left\{ \frac{1}{\left(\frac{1}{4}\right)^{\frac{1}{4}}} \right\} \frac{\varepsilon^{-\frac{3}{4}}}{\Gamma\left(\frac{1}{4}\right)} \exp\left\{ -\frac{4}{\varepsilon} \right\}, \tag{4.19}$$

which is a gamma density with $\alpha = \theta = \frac{1}{4}$.

The power of the test depends naturally on the distance between $\mathcal{H}_0$ and $\mathcal{H}_a$. Let $f_{\mathcal{H}_a}(\varepsilon)$ denotes the density function under the alternative hypothesis. Then to quantify this distance, we calculate

$$D(f_{\mathcal{H}_a}(\varepsilon)) = \int (f_E(\varepsilon, \theta_0) - f_{\mathcal{H}_a}(\varepsilon))^2 f_E(\varepsilon, \theta_0) d\varepsilon. \tag{4.20}$$

Table 4.5 summarizes the types of the distribution considered in this section and the distance $D(f_{\mathcal{H}_a}(\varepsilon))$ between the null and the alternative hypotheses.

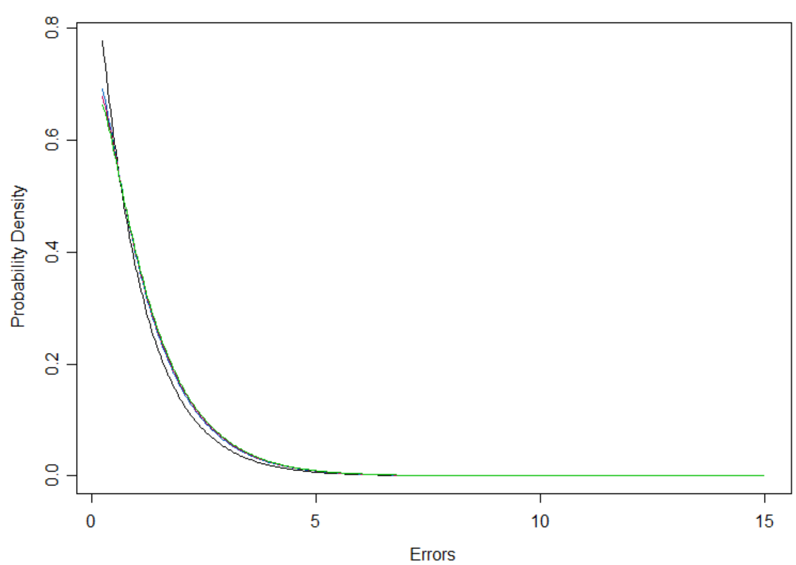
These distances are also shown graphically in Figures 4.1 and 4.2. In these figures, the black line represents the density under the null hypothesis, while other lines indicate the alternatives, which are further and further away from $\mathcal{H}_0$.

Table 4.5. Distance $D(f_{\mathcal{H}_a}(\varepsilon))$.

Testing Examples	Denotation	Densities	α	δ	β	$D(f_{\mathcal{H}_a}(\varepsilon))$
Set 1	E	$f_E(\varepsilon, \theta_0)$	1.0	1.0	1.0	·
	$G1$	$f_G(\varepsilon, \theta_1)$	1.3	1.0	1.0	0.0010
	$G2$	$f_G(\varepsilon, \theta_2)$	1.5	1.0	1.0	0.0025
	$G3$	$f_G(\varepsilon, \theta_3)$	1.7	1.0	1.0	0.0049

Testing Examples	Denotation	Densities	C_s	$D(f_{\mathcal{H}_a}(\varepsilon))$
Set 2	E	$f_E(\varepsilon, \theta_0)$	·	·
	SD_{C1}	$(1 - C_1) f_E(\varepsilon, \theta_1) + C_1\ \varphi(\varepsilon)$	0.15	0.0001
	SD_{C2}	$(1 - C_2) f_E(\varepsilon, \theta_2) + C_2\ \varphi(\varepsilon)$	0.17	0.0003
	SD_{C3}	$(1 - C_3) f_E(\varepsilon, \theta_1) + C_3\ \varphi(\varepsilon)$	0.19	0.0004

Figure 4.1. *E, G1, G2 and G3.*



Note that in the following tables, the quantities in the square brackets are computed as the following. Let $r = 1, \dots, 200$ represent the simulated replications,

$$Q = \{r | r \text{ is the replication, i.e. } 1, 2, \dots, 200, \text{ that is rejected by } L^*_{\varepsilon}\},$$

$$\widehat{Q} = \{r | r \text{ is the replication, i.e. } 1, 2, \dots, 200, \text{ that is rejected by } L^*_{\widehat{\varepsilon}}\}$$

and, therefore, $\widehat{Q} \bigcap Q$ is the set of those that are rejected by both L^*_{ε} and $L^*_{\widehat{\varepsilon}}$. The values in the square brackets are the quotient $\frac{R_1}{R_2}$ by which R_1 and R_2 are the numbers of elements in $\widehat{Q} \bigcap Q$ and Q, respectively. Hence, the closer to one the values in the square brackets are, the larger proportion of rejections using L^*_{ε}, which are also rejected by $L^*_{\widehat{\varepsilon}}$.

Let us focus first on the top panel of Tables 4.6 and 4.7. As expected, we are able to achieve higher power of the test for cases where the distances are relatively large. In all cases, the power of the test against the given set of fixed null hypotheses improves as T becomes large. Overall, the power values of L^*_{ε} look reasonable even with $\mathcal{H}_0$ and $\mathcal{H}_a$, which are quite close, and with such small numbers of observations.

Figure 4.2. *E*, SD_{C1}, SD_{C2} *and* SD_{C3}.

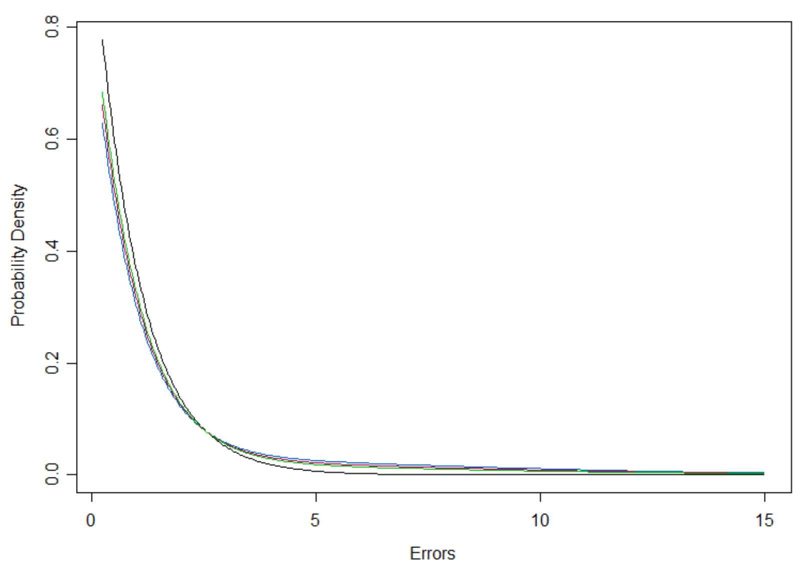


Table 4.6. Power of the test (Set 1).

Power of the Test L^*_ε against Fixed Alternatives

	$\alpha = 5\%$			$\alpha = 1\%$		
	$f_G(\cdot, \theta_1)$	$f_G(\cdot, \theta_2)$	$f_G(\cdot, \theta_3)$	$f_G(\cdot, \theta_1)$	$f_G(\cdot, \theta_2)$	$f_G(\cdot, \theta_3)$
$T = 176$	0.035	0.050	0.182	0.005	0.020	0.085
$T = 301$	0.120	0.345	0.475	0.015	0.115	0.195
$T = 501$	0.250	0.655	0.861	0.075	0.345	0.505

Power of the Test $L^*_{\hat{\varepsilon}}$ against Fixed Alternatives

	$\alpha = 5\%$			$\alpha = 1\%$		
	$f_G(\cdot, \theta_1)$	$f_G(\cdot, \theta_2)$	$f_G(\cdot, \theta_3)$	$f_G(\cdot, \theta_1)$	$f_G(\cdot, \theta_2)$	$f_G(\cdot, \theta_3)$
$T = 176$	0.060[0.580]	0.065[0.769]	0.175[0.914]	0.010[0.000]	0.015[0.670]	0.040[1.000]
$T = 301$	0.115[0.782]	0.390[0.871]	0.485[0.896]	0.005[1.000]	0.120[0.875]	0.160[0.875]
$T = 501$	0.190[1.000]	0.585[1.000]	0.790[1.000]	0.020[0.250]	0.155[0.903]	0.250[0.940]

Table 4.7. Power of the test (Set 2).

Power of the Test L^*_ε against Fixed Alternatives

	$\alpha = 5\%$			$\alpha = 1\%$		
	SD_{C1}	SD_{C2}	SD_{C3}	SD_{C1}	SD_{C2}	SD_{C3}
$T = 176$	0.115	0.180	0.205	0.030	0.075	0.085
$T = 301$	0.190	0.275	0.325	0.070	0.115	0.165
$T = 501$	0.335	0.520	0.630	0.160	0.280	0.400

Power of the Test $L^*_{\hat{\varepsilon}}$ against Fixed Alternatives

	$\alpha = 5\%$			$\alpha = 1\%$		
	SD_{C1}	SD_{C2}	SD_{C3}	SD_{C1}	SD_{C2}	SD_{C3}
$T = 176$	0.115[0.826]	0.180[0.805]	0.220[0.863]	0.025[0.600]	0.040[0.880]	0.040[0.750]
$T = 301$	0.245[0.775]	0.285[0.912]	0.395[0.757]	0.075[0.800]	0.140[0.714]	0.190[0.789]
$T = 501$	0.303[0.990]	0.505[0.970]	0.615[0.950]	0.090[0.777]	0.185[0.810]	0.275[0.909]

With regard to the bottom panel, let us illustrate, with a few examples, how these results can be interpreted. Observe that the results shown in the top and the bottom panels of Table 4.7 for $f_G(\cdot, \theta_3)$ at $T = 501$ are 0.861 and 0.790[1.000], respectively. Given that the total number of replications used in this study is $R = 200$, these results suggest that the number of rejection for L^*_ε is 176, i.e. 0.88×200, compared to 158 for $L^*_{\hat{\varepsilon}}$. Furthermore, the number in the square bracket suggests that, out of 172 rejections by L_{ε}^{*}, 158 have been correctly rejected by $L_{\widehat{\varepsilon}}^{*}$ leaving $28 + 14 = 42$, which the tests fail to reject.

Now, let us concentrate on cases by which $\alpha = 5\%$, i.e. columns 2 to 4, of the tables. In all cases, the power values of the tests have strong tendency to converge to 1 as T increases. The power values are reported to be quite close to 1 for both sets of examples, even with such a small number of observation of 501. This tendency is also what we can observe about the quotient $\frac{R_1}{R_2}$. Furthermore, the power values presented in the bottom panels are reasonably close to those of the top panels. Clearly, these results offer an experimental evidence in support of the asymptotic consistency of the test stated in Theorem 3.2 above.

4.2.3. Power of the Test against a Sequence of Local Alternatives

To study the power of the test against a sequence of local alternatives, the test in this subsection is constructed so that although globally ε has a gamma distribution with either monotonically increasing or decreasing hazard rate, it may deviate locally at some finite T by a ratio of some nonparametric density function, which converges to zero as $T \to \infty$.

Specifically, the null and alternative hypotheses in this example are constructed as

$$\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0) \; vs. \; \mathcal{H}_a : f_T(\varepsilon) = (1 - C_T) f_G(\varepsilon, \theta_1) + C_T \; \varphi(\varepsilon), \tag{4.21}$$

where $C_T = T^{-\frac{1}{2}} \sqrt{\log \log T}$. To introduce the above–mentioned local deviation, while ensuring that the alternative function in this case is a probability density and that it is fairly close to the density law under $\mathcal{H}_0$, we let

$$\varphi(\varepsilon) = \left(\frac{7}{10} \right) \frac{\varepsilon^{\frac{4}{10}}}{\Gamma(2)} \exp \left\{ -\varepsilon^{\frac{7}{10}} \right\}, \tag{4.22}$$

which is an inverted U–shaped hazard rate function generalized gamma density with $\alpha = 2$, $\delta = 0.7$ and $\theta = 1$. While the first column in Table 4.8 presents the values of the probability mixing parameters C_{176}, C_{301} and C_{501}, the second column displays the distances between the resulting mixing density functions and the null hypothesis. These distances are also demonstrated graphically in Figure 4.3, where the black line represent the density under $\mathcal{H}_0$.

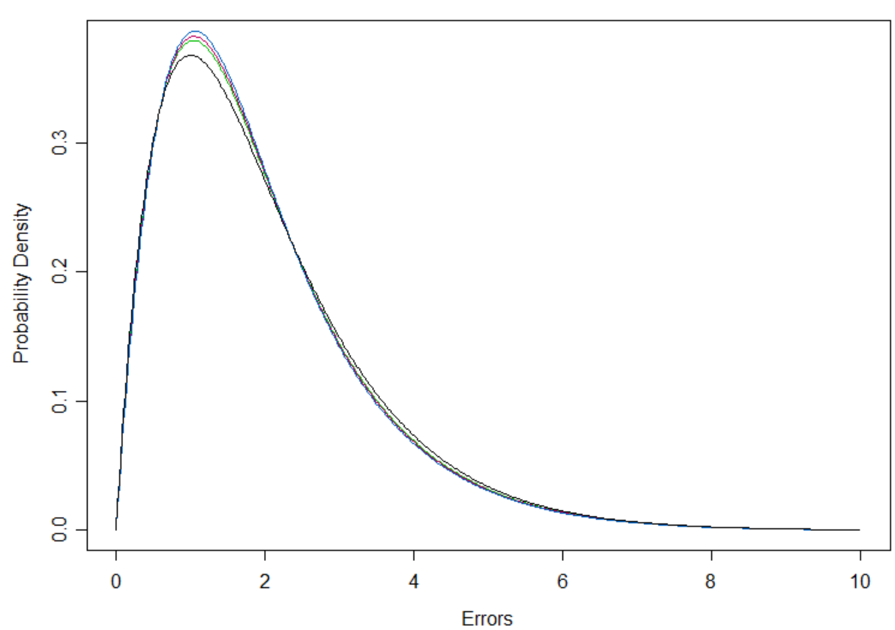
Table 4.8. Power of the Tests

Power of the Tests L^*_{ε} and $L^*_{\widehat{\varepsilon}}$ against a Sequence of Local Alternatives

			L^*_{ε}		$L^*_{\widehat{\varepsilon}}$	
C_T	$D(f_{\mathcal{H}_a}(\varepsilon))$	T	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
$C_{176} = 0.096$	0.00037	176	0.335	0.125	0.282[0.835]	0.115[0.800]
$C_{301} = 0.076$	0.00026	301	0.355	0.095	0.332[0.915]	0.085[0.842]
$C_{501} = 0.051$	0.00018	501	0.367	0.127	0.339[0.864]	0.135[0.920]

The results in the sixth and seventh columns in Table 4.8 can be interpreted in a similar manner as those of Tables 4.6 and 4.7 above. Using the distance in the second column as a criterion, it is clear that the power values of the tests with the alternative hypotheses associated with C_{176}, C_{301} and C_{501} should, at least, be comparable with those of SD_{C3} at $T = 176$, SD_{C2} at $T = 301$ and SD_{C1} at $T = 501$, respectively. The fact that the power values of 0.335, 0.355 and 0.367 in Table 4.8 are larger than 0.220, 0.285 and 0.303 of Table 4.7 is a convincing evidence in support of the asymptotic results in Theorem 3.3.

Figure 4.3. *Sequence of local alternatives.*



4.3. Boundary Bias and Bandwidth Selection in Nonparametric Kernel Testing

As mentioned earlier, the so–called boundary bias is the difficulty that may exist in nonparametric estimation using a fixed kernel due to the fact that duration has a support which is bounded from below. In an attempt to minimize the impact of such problem, Fernandes and Grammig (2005) follow a suggestion in the literature, such as Chen (2000), and establish an asymmetric kernel version of the test.

Fernandes and Gramming's so–called D-Test rests on measuring the distance

$$\Phi_f = \int_\varepsilon I(\varepsilon \in \mathcal{S})[f(\varepsilon, \theta) - f(\varepsilon)]^2 \, dF(\varepsilon), \tag{4.23}$$

where $\int_\varepsilon$ denotes the integral over the support of the density function ε and $I(\cdot)$ is the indicator function. The D-Test gauges the discrepancy between the parametric and nonparametric estimates of the density function, which gives light to the following functional of interest

$$\Phi_{\widetilde{f}} = \int_\varepsilon I(\varepsilon \in \mathcal{S})[f(\varepsilon, \theta_{\widetilde{f}}) - \widetilde{f}(\varepsilon)]^2 \, dF_T(\varepsilon), \tag{4.24}$$

where $\theta_{\widetilde{f}}$ and $\widetilde{f}(\cdot)$ denote pointwise consistent estimates of the true parameter θ_0 and density $f(\cdot)$, respectively. Fernandes and Grammig (2005) estimate the density function using the gamma kernel

$$k_{\frac{x}{b_T+1}, b_T}(u) = \frac{u^{\frac{x}{b_T}} \exp\left(-\frac{u}{b_T}\right)}{\Gamma\left(\frac{x}{b_T+1}\right) b_T^{\frac{x}{b_T+1}}} I(u \geq 0) \tag{4.25}$$

with bandwidth b_T, while show that

$$\widetilde{L}_{FG}(b_T) = \frac{T b_T^{\frac{1}{4}} \Phi_{\widetilde{f}} - b_T^{\frac{1}{4}} \widetilde{\delta}}{\widetilde{\sigma}} \xrightarrow[D]{} N(0, 1),$$

where $\widetilde{\delta}$ and $\widetilde{\sigma}$ are consistent estimates of

$$\delta = \frac{1}{\sqrt{\pi}} E\left[I(\varepsilon \in \mathcal{S}) \varepsilon^{-\frac{1}{2}} f(\varepsilon)\right] \text{ and } \sigma^2 = \frac{1}{\sqrt{\pi}} E\left[I(\varepsilon \in \mathcal{S}) \varepsilon^{-\frac{1}{2}} f^3(\varepsilon)\right],$$

respectively. In practice, the bandwidth of the test is selected using an under–smoothing adjusted version of the rule–of–thumb bandwidth, i.e.

$$\widehat{b}_T = \frac{1}{\log T} \left\{ \frac{\widehat{\lambda}}{4} \right\}^{-\frac{1}{5}} (2 - \widehat{\lambda})^{-\frac{4}{5}} T^{-\frac{4}{9}},$$

where $\widehat{\lambda}$ is some consistent estimator of the exponential parameter λ.

We compare in this section the size and power properties of $\widehat{\widehat{L}}_{FG}(\widehat{b}_T)$ computed using $\widehat{\Phi}_{\widetilde{f}}$, which is the sample analog of (4.24), and L^*_ε. The comparison is done under the following null and alternative hypotheses

$$\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0) \;\; vs. \;\; \mathcal{H}_a : f(\varepsilon) = \{0.95 \times f_G(\varepsilon, \theta_1)\} + \{0.05 \times \; f_W(\varepsilon, \theta_1)\}. \quad (4.26)$$

Specifically, $f_G(\varepsilon, \theta_0)$ and $f_G(\varepsilon, \theta_1)$ are the density functions of a gamma distribution with $\alpha = 2$ and $\theta = 2$, and $\alpha = 2$ and $\theta = 3$, respectively. Furthermore, $f_W(\varepsilon, \theta_1)$ is the density function of a weibull distribution with $\theta = 3$ and $\delta = 1$.

In order to focus the discussion in this section specifically on the comparison of the two tests, we only concentrate on the case where a random sample is generated from a known density under either the null or the alternative hypothesis. Three different sizes of sample, namely $T = 300, 500$ and 700 are considered. The corresponding simulated critical values at the α-level are found by using the simulation scheme proposed in Section 3, which is implemented at $M = 1,000$. The number of simulations used in producing the proceeding tables in this section is 200. The detailed results at the 1%, 5% and 10% significance levels are given in Table 4.9.

Table 4.9. Simulated size and power values at 1%, 5% and 10% significance levels.

Sample size	Null hypothesis is true.		Null hypothesis is false.	
T	$\widehat{\widehat{L}}_{FG}(\widehat{b}_T)$	L^*_ε	$\widehat{\widehat{L}}_{FG}(\widehat{b}_T)$	L^*_ε
1% significance level				
100	0.005	0.005	0.065	0.185
300	0.015	0.015	0.065	0.715
500	0.020	0.015	0.100	0.950
700	0.010	0.010	0.080	0.995
5% significance level				
100	0.040	0.045	0.145	0.380
300	0.045	0.055	0.175	0.830
500	0.005	0.055	0.205	0.995
700	0.040	0.050	0.245	1.000
10% significance level				
100	0.085	0.100	0.230	0.495
300	0.095	0.090	0.260	0.910
500	0.080	0.115	0.315	1.000
700	0.070	0.095	0.320	1.000

Table 4.9 reports comprehensive simulation results for both the size and power values of the two test statistics. It is clear in the third column of each panel of the table that the sizes of L_{ε}^{*} are more consistent across both significance levels and number of observations. The gamma–kernel–based test, $\widehat{\widehat{L}}_{FG}(\widehat{b}_T)$, seems to perform relatively poorly in this aspect. Furthermore, the power values of L_{ε}^{*} reported in the forth column are always greater than those of $\widehat{\widehat{L}}_{FG}(\widehat{b}_T)$ in the third column. It seems that $\widehat{\widehat{L}}_{FG}(\widehat{b}_T)$ achieves lower power values in all categories.

5. Econometric Analysis of Price Duration Process

Having defined the inter–event waiting times as financial durations, we are able to classify these durations further according to their events of interest. Some of the most commonly studied duration processes in high frequency data literature are trade and quote durations which are defined as the time between two consecutive trade and quote arrivals, respectively. More recently, much attention has also been paid on the econometric modeling of price durations, which correspond to the time between cumulative absolute price changes of a given size. The econometric modeling of price durations is important for at least three reasons, namely (i) there is a direct relationship between the price intensity and volatility as pointed out by Engle and Russell (1998), (ii) the behavior of price durations has important implications for option pricing as shown by Prigent et al. (2000) and (iii) price duration process can be used to empirically test microstructure theories as demonstrated by Bauwens et al. (2004) and Engle and Russell (1998).

The empirical analysis in this paper applies the above explained two–staged estimation and testing to study the price duration process at the NYSE and the ASX. There are two sets of data used in this analysis. The first dataset is the IBM data used in Engle and Russell (1998). A total of 60,328 transactions were recorded for IBM over the three months of trading on the consolidated market from November 1990 through January 1991. As per the seminal paper, two days from the three months were deleted. A halt occurred on 23rd November and a more than one hour opening delay occurred on 27th December. The second dataset is that of AMP stock, which is listed on the Australian Stock Exchange. The dataset is tick–by–tick data for the period of April to June 2000.

The first half hour of the trading days, i.e. trades before 10:00am, are omitted. This is to avoid modeling the market which is characterized by a call auction followed by heavy activity. The dynamics are likely to be quite different over this period. Furthermore, the call auction transactions are not recorded at the same time each morning. In addition, all trades after 4:00pm are also omitted.

To construct the price duration processes, the so–called dependent thinning is performed. In essence only the points at which the price has changed significantly since the occurrence of the last price change is kept. In order to to minimize the effects of errant quotes two consecutive points were required to have changed more than a threshold value, c, since the last price change. Engle and Russell (1997) provide a more formal explanation of the thinning method as follows:

(i) Retain point 1.

(ii) Retain point $s > 1$ if $|p_s - p_j| > c$ and $|p_{s+1} - p_j| > c$ where j is the index of the most recent retained point and the constant c represents a threshold value.

It is widely documented in the high–frequency data literature, for example Giot (2000), that 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. In the current paper, we assume that the stationary price duration series, which to be modeled in the SEMI–ACD procedure, can be computed as

$$x_t = \frac{\nu_t}{\phi(i_{t-1})} = \psi_t \varepsilon_t, \tag{5.1}$$

where ν_t denotes the observed price duration as constructed earlier and $\phi(i_{t-1})$ denotes an intraday diurnal factor.

Table 5.1. Descriptive Statistics

Descriptive Statistics	ν_{IBM}	$\tilde{x}_{IBM}$	$\widehat{\varepsilon}_{IBM}$	ν_{AMP}	$\tilde{x}_{AMP}$	$\widehat{\varepsilon}_{AMP}$
Mean	892.58	1.36	1.02	900.66	0.93	1.01
Standard Deviation	1258.89	2.09	1.58	1478.88	1.48	1.59
Kurtosis	11.26	14.82	15.30	19.94	21.70	21.37
Skewness	2.86	3.29	3.35	3.82	3.92	3.93
Minimum	1.00	0.00	0.00	1.00	0.00	0.00
Maximum	10609.00	19.12	14.45	15082.00	15.92	15.52
Ljung–Box[10]	42.50	34.20	23.44	91.60	79.18	24.55
	(0.000)	(0.000)	(0.012)	(0.000)	(0.000)	(0.010)

The second and the fifth columns of of Table 5.1 present the descriptive statistics of IBM and AMP observed price duration, respectively. Note that the variables in the remaining columns of the table will be considered at a later stage. While, the average price durations for the IBM and AMP samples are 893 and 900 seconds, the maximum are 10,609 and 15,082 seconds, respectively. Minimum price duration for both IBM and AMP is 1 second. The Ljung–Box test values of 42.50 and 91.60 indicate strong clustering behavior and autocorrelation in both IBM and AMP price duration series.

Following a similar modeling procedure in Wongsaart et al. (2011), in this paper we estimate the diurnal factor $\phi(i_{t-1})$ of the calendar time i_{t-1} at which the t-th duration begins using the kernel regression smoothing technique with the smoother defined as

$$\widehat{\phi}_h\left(i_{t-1}\right) = \sum_{v=1}^{N} W_{v,h}(i_{t-1})\nu_v, \tag{5.2}$$

where $W_{v,h}(y) = \frac{K_h(y-i_{t-1})}{\sum_{t=2}^{T} K_h(y-i_{t-1})}$ is a kernel weight function. In our calculation, an asymptotically optimal bandwidth parameter is selected using the leave–one–out cross validation selection criterion such that $H_T = \left\{h = h_{\max}a^k : h \geq h_{\min}, k = 0, 1, 2, \ldots\right\}$, where $0 < h_{\min} < h_{\max}$ and $0 < a < 1$. Letting J_T denotes the number of elements of H_T, we define $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$.

Figure 5.1. *Expected price duration on hour of day for AMP (left panel) and IBM (right panel).*

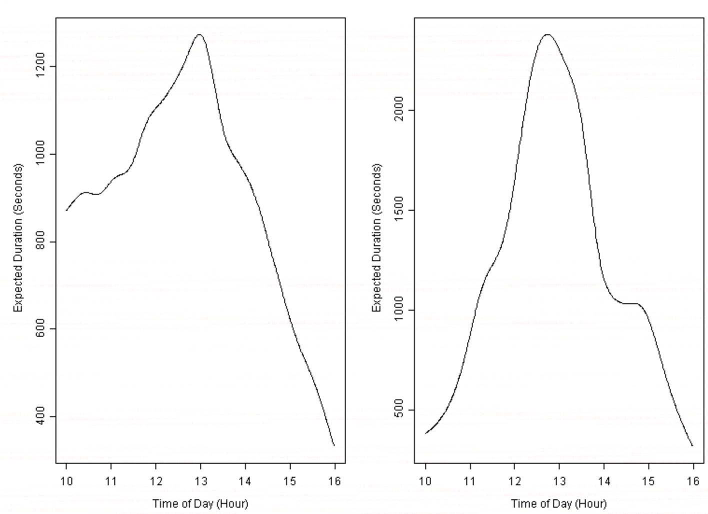


Figure 5.1 presents the kernel estimates of the diurnal factors associated with the IBM and AMP price duration processes. As expected, the price durations are shortest in the

morning and just prior to the close with a noticeable lull between during lunch hours. These results are consistent with those found in existing literature; see, for example, Engle and Russell (1998).

The next step is to model the ratio of the actual to the fitted value $\widetilde{x}_t = \frac{\nu_t}{\widehat{\phi}_h(i_{t-1})}$ as a SEMI–ACD(1,1) model of the diurnally adjusted series of price durations. However, before doing so, let us also check to see if diurnal adjustment alone is able to take care of serial correlation and duration clustering. The Ljung–Box statistic reported in the third and the sixth columns of Table 5.1 suggests that the diurnal adjusted price duration series of both IBM and AMP still strongly exhibit these peculiar time series features. Hence, the use of the SEMI–ACD procedure to model the stochastic component of price duration processes is essential.

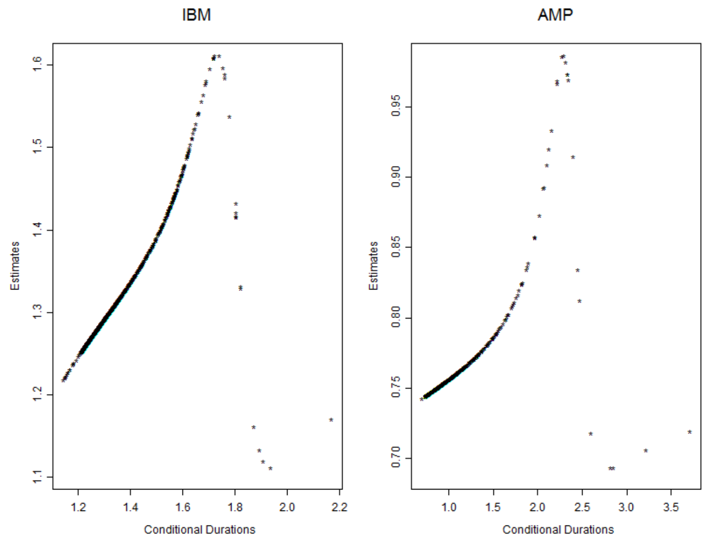
To this end, a number of previous studies in the field of nonparametric kernel estimation have suggested that the choice of the kernel function is much less critical than the bandwidth choice. To estimate the SEMI–ACD model, we employ the quartic kernel function

$$K(u) = \begin{cases} \left(\frac{15}{16}\right)(1-u^2)^2 & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

An estimation–based optimal bandwidths for each of the iteration step is selected using an adaptive cross validation algorithm as discussed in Section 4.1. Finally, we select $m^* = 7$ by using a similar procedure discussed in detail in Wongsaart et al. (2011).

Hereafter, we let "IBM–ACD" and "AMP–ACD" abbreviate the SEMI–ACD(1,1) models of IBM and AMP, respectively. In the IBM–ACD and AMP–ACD models, the kernel WLSE estimates of the unknown parameter γ are 0.0363(0.1459) and 0.1852(0.1585), respectively. (The values in the parentheses are the standard errors.) Furthermore, the left and right panels of Figure 5.2 present the empirical estimate of g in the IBM–ACD and the AMP–ACD models, respectively. Even though the two curves look similar in shape, the figure suggests a much stronger influence of previous durations on the dynamics of the duration process for AMP. On the contrary, the IBM–ACD suggests that the dynamics of the process is mainly driven by the lag value of conditional duration. Furthermore, in obtaining these results, we also find that the estimation–based optimal bandwidths selected for the SEMI–ACD algorithm are close to 0.3 for IBM–ACD, while they are about 1.25 for the case of AMP–ACD.

Figure 5.2. *Empirical estimate of the unknown real valued function $g(\cdot)$.*



To proceed with the hypothesis testing, the empirical estimates of the standardized durations must now be computed based on the formula

$$\widehat{\varepsilon}_{t,m^*} = \frac{\nu_t}{\widehat{\phi}_h(i_{t-1})\widehat{\psi}_{t,m^*}}. \tag{5.4}$$

The descriptive statistics of the series is presented in Table 5.1 above.

There are numerous suggestions in the duration literature on how the baseline hazard for the price durations can be empirically estimated. We consider in this paper an alternative approach, which is to (1) estimate the density of the empirical standardized durations using kernel density estimation, (2) compute the associated survival function and (3) take the quotient of the two to obtain the baseline hazard. The survival function of ε is the function S_ε defined by $S_\varepsilon(e) = Pr(\varepsilon > e)$ for all e. Let us define $\widehat{f}_\varepsilon(e)$ as a nonparametric kernel density estimate of the form

$$\widehat{f}_\varepsilon(e) = \frac{1}{Th}\sum_{t=1}^{T} k\left(\frac{e_t - e}{h}\right), \tag{5.5}$$

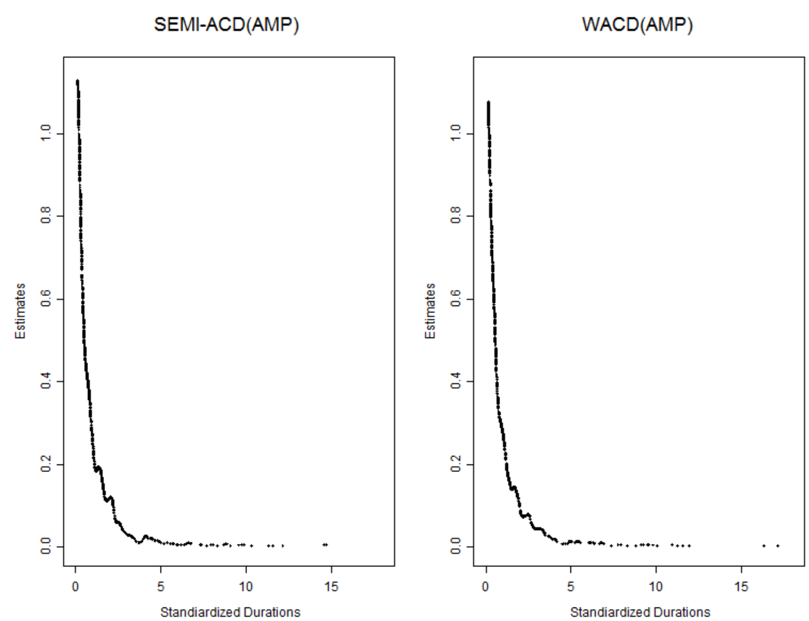
where h is the bandwidth parameter. If $k(z)$ is the normal kernel function, Wongsaart et al. (2011) suggest that we estimate $S_\varepsilon(e)$ using

$$\widehat{S}_\varepsilon(e) = \frac{1}{T}\sum_{t=1}^{T} \Phi\left(\frac{e_t - e}{h}\right) \tag{5.6}$$

where $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp\left\{\frac{-u^2}{2}\right\} du$. In order to use the formulae in (5.5) and (5.6), we only have to replace e_t by $\widehat{\varepsilon}_{t,m^*}$.

Figures 5.3 and 5.4 present the kernel density estimates of the density functions of the standardized durations for IBM and AMP, respectively. In these cases, the bandwidth parameter used is an estimation–based optimal bandwidth that is selected based on the unbiased cross validation criterion; see, for example, Li and Racine (2007). The figures illustrate graphically the impacts of the new semiparametric regression on the subsequent density estimation by comparing the kernel density estimate based on the SEMI–ACD to that of Engle and Russell's Weibull ACD (WACD) model. There are clear differences in the shape of the estimates for both IBM and AMP. Below, we employ our newly developed testing procedure to investigate, if further evidence can be established in support of such graphical findings.

Figure 5.3. *Kernel density estimate of the density function for IBM: SEMI–ACD vs WACD.*



However, unlike in the above finite sample studies, γ and g are both unknown in applications. Hence, in order to apply our testing procedure, we propose the following steps for computing the p-values of $L^*_{\widehat{\varepsilon}}$:

(5.1) Compute $\widehat{\psi}_{t,m^*}$ and generate $\{\widehat{\varepsilon}^*_t\}$, which is a sequence of i.i.d. bootstrap re–samples generated under the null hypothesis.

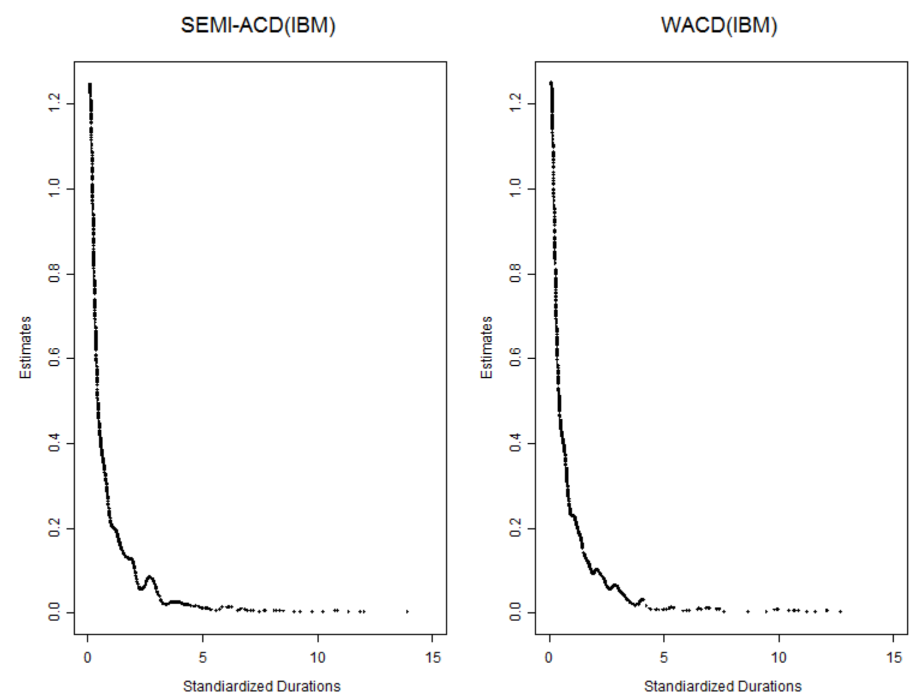
(5.2) Compute $\widehat{x}_t^* = \frac{\widehat{\psi}_{t,m^*}}{\widehat{\varepsilon}_t^*}$ and the corresponding $L_{\widehat{\varepsilon}}^*$ based on

$$\widehat{x}_t^* = \widehat{\gamma}_{\widehat{\psi}}(h)x_{t-1} + \widehat{g}_h(\widehat{\psi}_{t,m^*}) + \widehat{\eta}_t^*,$$

where $\widehat{\eta}_t^* = \widehat{\psi}_{t,m^*}(\widehat{\varepsilon}_t^* - 1)$.

(5.3) Repeat the proceeding steps M times in order to produce M versions of $L_{\widehat{\varepsilon}}^*$, i.e. $L_{\widehat{\varepsilon},m}^*$ for $m = 1, 2, \cdots, M$. Find the bootstrap distribution of $L_{\widehat{\varepsilon},m}^*$ and then compute the proportion in which $L_{\widehat{\varepsilon}}^* < L_{\widehat{\varepsilon},m}^*$. This proportion is then a simulated p-value of $L_{\widehat{\varepsilon}}^*$.

Figure 5.4. *Kernel density estimate of the density function for AMP: SEMI–ACD vs WACD.*



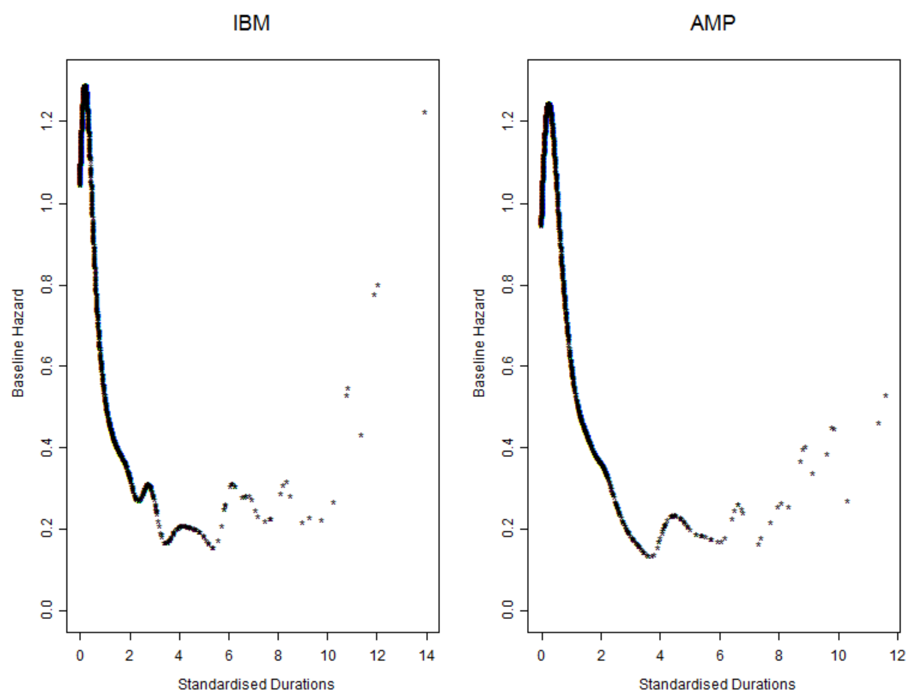
There is also a few other practical issues that must be addressed prior to performing our testing procedure. The first issue concerns the set of parameters to be used with the distribution under the null hypothesis in step 5.1. Our experience suggests that the resulting p-values can be vastly different from one type to another. In this paper, we first obtain the maximum likelihood estimates of the parameters of the density function under the null hypothesis, then use these to generate $\{\widehat{\varepsilon}_t^*\}$.

Secondly, the use of the maximized version of the test statistic suggests that selection H_T can have significant impacts on the final conclusions of the results. For both estimation and testing, a specific formulation of H_T employed here is a geometric grid of the form

$$H_T = \{h = h_{\max} a^k : h \geq h_{\min},\ k = 0, 1, 2, \ldots\}, \tag{5.7}$$

where $0 < h_{\min} < h_{\max}$ and $0 < a < 1$. In this case, $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$. More detailed conditions on $h_{\max}$ and $h_{\min}$ are given in Appendix A. In order to perform hypothesis testings, we choose $a = 0.25$ and $b = 0.4$.

Figure 5.5. *Empirical estimate of baseline hazard for price durations.*



The third issue involves choices of parametric marginal density functions to be tested. Figure 5.5 gives the empirical estimates of the baseline hazards of the two sets of data, namely IBM and AMP. Both panels of the figure display curves that are essentially downward slopping. Therefore, an exponential distribution, which implies a constant hazard function, can be safely excluded for both cases. In fact the monotonic shape of the curves in Figure 5.5 is consistent with that of the hazard functions of the gamma and weibull distributions. Nonetheless, there are some aspects of the figures which suggest that overall the curves might be U–shaped and are therefore consistent with a hazard function of a generalized gamma distribution. In order to provide further evidence either for or against such findings, the parametric density functions that we will concentrate our testing on are those of the gamma, weibull and the generalized gamma distributions.

Finally, it is well known that if a probability density function has bounded support, kernel density estimates often overspill the boundaries and are consequently especially biased at and near these edges (Jones (1993)). Even though our earlier experimental results perform well in finite sample, in order to minimize the potential impact of such the problem and given the fact, as can be evidenced in Figures 5.3 and 5.4, that only a small proportion of the sample is near zero, the test statistics of all testings considered here are computed as

$$\widehat{N}_{T,\widehat{\varepsilon}}(h) = h \sum_{t=1}^{T} \{\widehat{f}(\widehat{\varepsilon}_{t,m^*}) - \widetilde{f}(\widehat{\varepsilon}_{t,m^*}, \widetilde{\theta})\}^2 \; \omega(\widehat{\varepsilon}_{t,m^*}), \tag{5.8}$$

where $\omega(\cdot)$ is a weight function such that $\omega(y) = \begin{cases} 1 & \text{if } |y| \leq 0.05 \\ 0 & \text{otherwise} \end{cases}$.

Hereafter, let us denote the standardized duration process of the WACD model and its WACD based estimate by ξ_t and by $\widehat{\xi}_t$, respectively. Our testing strategies are the following:

Test 1: We test the null $\mathcal{H}_{01} : \exists \theta_0 \in \Theta$ such that $f_W(\xi, \theta_0) = f(\xi)$ against the alternative hypothesis that there is no such $\theta_0 \in \Theta$ as a mean to assess the goodness of fit of the weibull distribution assumption.

Test 2: For the sake comparison, we test the following hypotheses:

(i) $\mathcal{H}_{021} : \exists \theta_0 \in \Theta$ such that $f(\varepsilon) = f_W(\varepsilon, \theta_0)$ against the alternative hypothesis that there is no such $\theta_0 \in \Theta$.

(ii) $\mathcal{H}_{022} : \exists \theta_0 \in \Theta$ such that $f(\varepsilon) = f_G(\varepsilon, \theta_0)$ against the alternative hypothesis that there is no such $\theta_0 \in \Theta$.

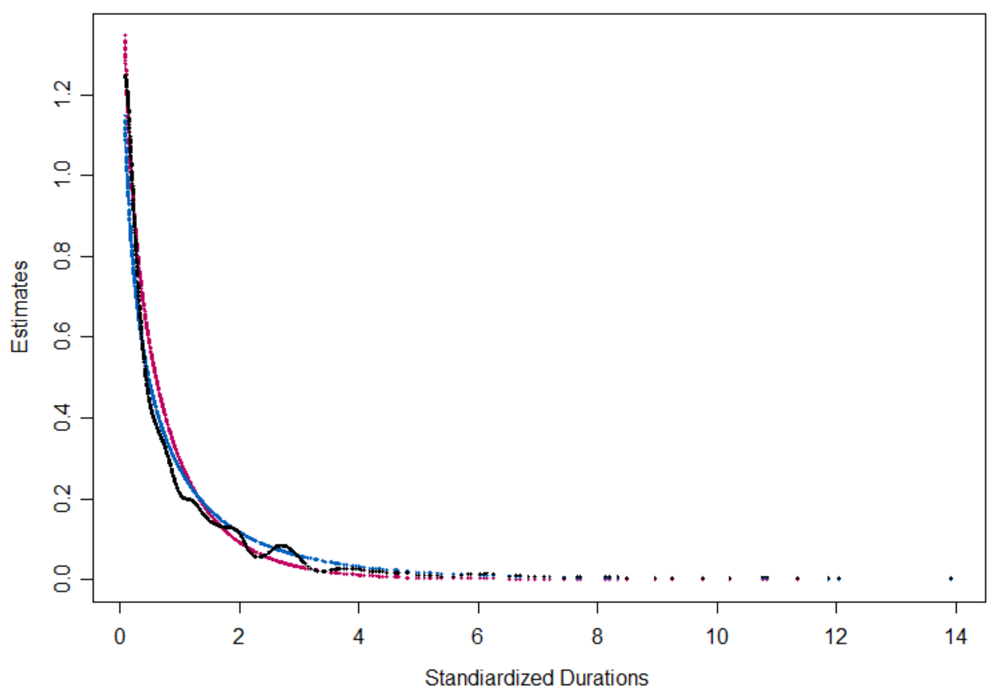
Table 5.2 reports our test results. The table shows strong evidence of a misspecification of the WACD model for both IBM and AMP. Such results are consistent with the findings in Engle and Russell (1998). With regard to AMP, the nonlinearity in the SEMI–ACD model is able to affect the testing result such that the null hypothesis of the weibull distribution is no longer rejected at the 5% significance level; see Test 2(i). Nonetheless, this is not the case for IBM. Even with the SEMI–ACD model, the null hypothesis of the weibull distribution is still rejected at the 5% significance level. Furthermore, the null hypothesis of the gamma distribution also faces a similar rejection for both IBM and AMP. Therefore, we only are able to find some empirical evidence based on the SEMI–ACD specification in support of the gamma distribution for the case of AMP. Some further analysis is clearly required for IBM.

Table 5.2. Hypothesis testing on the parametric density function of the standardized duration.

Testing Strategy	Distribution	IBM/AMP	p-values
Test 1:	Weibull	IBM	0.024
		AMP	0.000
Test 2(i):	Weibull	IBM	0.019
		AMP	0.163
Test 2(ii):	Gamma	IBM	0.021
		AMP	0.000

Note: p-values are based on the empirical distribution of the test statistic stemming from 200 artificial bootstrap samples.

Figure 5.6. *Kernel density estimate of the density function.*



The black–dotted line in Figure 5.6 presents the nonparametric estimate of the density of the standardized duration, while the pink–dotted and blue–dotted lines display density functions of the gamma distribution (with $\alpha = 0.75$ and $\theta = 1$) and the weibull distribution (with $\theta = 1$ and $\delta = 0.75$), respectively. While the gamma distribution fits the empirical estimate better for the values of the standardized durations below 1, the figure also shows that the weibull distribution fits better for larger values. This graphical evidence suggests that a mixture of the weibull and gamma distributions might be appropriate. This leads us to the third testing strategy.

Test 3: We apply our test to testing

$$\mathcal{H}_{03}: \ f(\varepsilon) = (1-c_0) f_W(\varepsilon, \theta_0) + c_0 f_G(\varepsilon, \theta_0) \text{ versus } \mathcal{H}_{13}: \ f(\varepsilon) = (1-c_1) f_W(\varepsilon, \theta_1) + c_1 \varphi(\varepsilon),$$

where φ denotes a nonparametric density.

For the sake of comparison, Figure 5.7 presents the gamma density (with $\alpha = 0.65$ and $\theta = 1$), the weibull density (with $\theta = 1$ and $\delta = 0.70$) and the weibull–gamma mixture density (with the mixture parameter $c = 0.61$) of standardized durations, where α, δ and c are estimated using a maximum likelihood method.

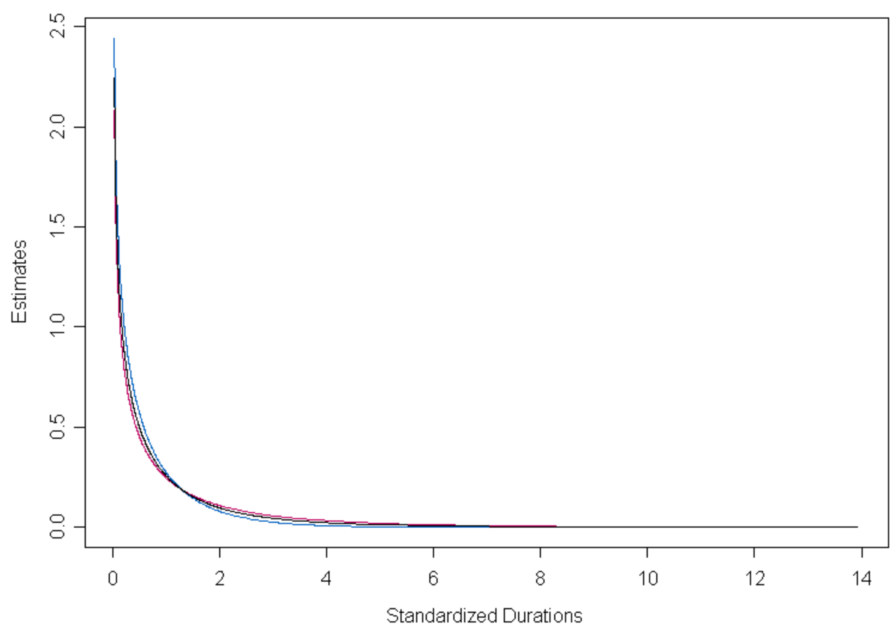
Table 5.3. Hypothesis testing on the parametric density function of the standardized duration.

Testing Strategy	Distribution	IBM/AMP	p-values
Test 3:	Mixture Weibull	IBM	0.44
	Generalized Gamma	IBM	0.13
		AMP	0.65

Note: p-values are based on the empirical distribution of the test statistic stemming from 200 artificial bootstrap samples. The test of the null hypothesis of the generalized gamma distribution was implemented in the Generalized Additive Models for Location Scale and Shape (GAMLSS) package in R.

With regard to the scale parameter, we set $\theta = 1$. In order to implement the test, the remaining parameters, namely α, δ and the mixture parameters, are estimated using maximum likelihood estimation. Unlike in Table 5.2, for IBM, the results in Table 5.3 show that the null hypothesis of a mixture weibull–gamma distribution is not rejected at the usual 5% significance level. Furthermore, the fact that the three distributions, namely weibull, gamma and the mixture weibull–gamma, are all nested within the three–parameter generalized gamma distribution, i.e. $f_{GG}(\varepsilon; \alpha, \theta, \delta)$, suggests that it should also be interesting to test the null hypothesis of such a generalized distribution. The results in Table 5.3 show, not surprisingly, that the null hypothesis of the three–parameter generalized gamma distribution is not rejected at the usual 5% significance level for both IBM and AMP.

Figure 5.7. *Gamma density, weibull density and weibull–gamma mixture density.*



6. Conclusions

We have presented in this paper a two–step semiparametric kernel procedure to test the parametric density function of financial durations. We have also illustrated that the nonparametric kernel testing introduced here also performs well when implemented to test hypotheses based on stationary random variables in general. The procedure introduced in this paper deviates significantly from that of Gao and King (2004). The most important feature of our hypothesis testing in such environment is the fact that there exists a latency problem that arises due to unobservability of the standardized duration. In this paper, we have shown theoretically and experimentally the asymptotic costlessness of our testing procedure under such the circumstance. Our experimental analysis also sheds further light on the issues of boundary bias and bandwidth selection. While both are the most common problems facing nonparametric kernel testing, existing literature, such as Grammig and Maurer (2000), attempt to address only the former. In this paper, although we concentrate mainly on the later, we have also provided some experimental evidence that suggests that our testing procedure performs better in the finite sample evaluation than its asymmetric kernel test counterpart. Finally, we have demonstrated how our two–step procedure can be used to model and test the price duration process in practice using data sets from the NYSE and the ASX. An important finding obtained from such an exercise is the fact that the semiparametric functional form specification of the conditional duration is able to influence the outcome of our hypothesis testing of parametric density function. Furthermore, we have concluded that a semiparametric mixture density can be a useful model for waiting–time in finance.

Appendices

Appendices A and B introduce a number of assumptions, which are required to establish that the relevant asymptotic results established in Gao and King (2004) can be extended to deal with the case where the unobserved $\{\varepsilon_t\}$ is replaced by $\{\widehat{\varepsilon}_{t,m^*}\}$. The proofs of (3.21) and Theorems 3.1–3.3 are given in Appendix C of the supplementary document.

Appendix A

Assumption A.1. *(i) Assume that the processes $\{x_t\}$ are strictly stationary and α-mixing with the mixing coefficient $\alpha(t) \leq C_\alpha \alpha^t$ defined by*

$$\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^2, B \in \Omega_{s+t}^\infty\}$$

for all $s, t \geq 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants and Ω_i^j denotes the σ-field generated by $\{x_t : i \leq t \leq j\}$.

(ii) k is a symmetric probability kernel function and it is four–time differentiable on $R^1 = (-\infty, \infty)$ with $\int \left|k^{(r)}(u)\right| du < \infty$ for $r = 1, 2$. In addition, k has an absolutely integrable Fourier transform with $\int_{-\infty}^{\infty} u^2 k\left(u\right) du < \infty$ and $\int_{-\infty}^{\infty} k^2(u) du < \infty$.

Assumption A.2. *(i) The parameter space $\Theta \subset R^q$ is compact. In a neighborhood of the true parameter θ_0, $f(\varepsilon, \theta)$ is twice continuously differentiable in θ; $E[(\partial f(\varepsilon, \theta)/\partial\theta)(\partial f(\varepsilon, \theta)/\partial\theta)^\tau]$ is full rank. In addition, assume that $G(\varepsilon)$ is a positive and integrable function with $E[G(\varepsilon_t)] < \infty$ uniformly in $t \geq 1$ such that $\sup_{\theta\in\Theta} |f(\varepsilon_t, \theta)|^2 \leq G(\varepsilon_t)$ and $\sup_{\theta\in\Theta} ||\nabla_\theta^j f(\varepsilon_t, \theta)||^2 \leq G(\varepsilon_t)$ for $j = 1, 2, 3$, where for $B = \{b_{ij}\}_{1\leq i,j\leq q}$, $||B^2||^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij}^2$.*

(ii) Assume that $\widetilde{\theta}$ is a $\sqrt{T}$-consistent estimator of θ_0.

Assumption A.3. *(i) Assume that the first three derivatives of $f(\varepsilon)$ are continuous on S_ω and that $f(\varepsilon) > c_f > 0$ on the interior of S_ω for some $c_f > 0$. In addition, $E\left[\left|f^{(r)}(\varepsilon_1)\right|\right] < \infty$ for $1 \leq r \leq 3$.*

(ii) The initial random variable ε_0 is distributed as $f(\varepsilon)$.

Assumption A.4. *The bandwidth parameter h satisfies the following conditions*

$$\lim_{T\to\infty} h = 0, \qquad \lim_{T\to\infty} Th^2 = \infty \quad \textit{and} \quad \lim\sup_{T\to\infty} Th^5 < \infty.$$

Assumption A.5. *The parameter set Θ is an open subset of R^q for some $q \geq 1$. The parametric family $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ satisfies the following conditions.*

(i) For each $\theta \in \Theta$, $f(\varepsilon, \theta)$ is continuous with respect to $x \in D$.

(ii) Assume that there is a finite $C_1 > 0$ such that for every $\varepsilon > 0$

$$\inf_{\theta,\theta' \in \Theta: ||\theta-\theta'|| \geq \varepsilon} [f(\varepsilon_1, \theta) - f(\varepsilon_1, \theta')]^2 \geq C_1 \varepsilon^2$$

holds with probability one (almost surely).

Assumption A.6. *(i) Let $\mathcal{H}_0$ be true. Then $\theta_0 \in \Theta$ and*

$$\lim_{T\to\infty} P(\sqrt{T}||\widetilde{\theta} - \theta_0|| > C_L) < \epsilon$$

for any $\epsilon > 0$ and all sufficiently large C_L.

(ii) Let $\mathcal{H}_0$ be false. Then there is a $\theta^* \in \Theta$ such that

$$\lim_{T\to\infty} P(\sqrt{T}||\widetilde{\theta} - \theta^*|| > C_L) < \epsilon$$

for any $\epsilon > 0$ and all sufficiently large C_L.

Assumption A.7. *(i) Let H_T be specified in (3.17) with*

$$c_{\min} T^{-\gamma} = h_{\min} < h_{\max} = c_{\max} \left(\text{loglog}(T)\right)^{-1},$$

where γ, $c_{\min}$ and $c_{\max}$ are some constants satisfying $0 < \gamma < 1$ and $0 < c_{\min}, c_{\max} < \infty$.

(ii) Suppose that $\Delta_T(\varepsilon)$ is continuous in ε and satisfies $\int_{-\infty}^{\infty} \Delta_T(\varepsilon) d\varepsilon = 0$ for all $T \geq 1$.

Appendix B

The list below shows the necessary assumptions for the establishment and the proofs of Theorems 3.1 to 3.3.

Assumption B.1. *Assume that function* g *on the real line satisfies the following Lipschitz type contraction property:*

$$|g(y) - g(x)| \le \varphi(x)|y - x| \tag{B.1}$$

for each given x *and* $y \in S_\omega$*, where* S_ω *is the compact support of the weight function* $\omega(\cdot)$ *as assumed in Assumption B.3 below and* $\varphi(x)$ *is a nonnegative measurable function such that*

$$\max_{i\ge 1} E\left[\varphi^2(\psi_i)|\psi_{i-1}, \cdots, \psi_1\right] \le G^2$$

with probability one for some $0 < G < 1$.

Assumption B.2. *(i) Suppose that the error process* $\{\varepsilon_t\}$ *and the conditional duration* $\{\psi_t\}$ *are both strictly stationary and* α*–mixing with mixing coefficients* $\alpha_\varepsilon(T)$ *and* $\alpha_\psi(T)$ *satisfying*

$$\alpha_\varepsilon(T) \le C_\varepsilon q_\varepsilon^T \quad and \quad \alpha_\psi(T) \le C_\psi q_\psi^T, \tag{B.2}$$

respectively, where $0 < C_\varepsilon,\ C_\psi < \infty$ *and* $0 < q_\varepsilon, q_\psi < 1$.

(ii) Let $\{\varepsilon_t\}$ *satisfy* $E[\varepsilon_1] = 1$ *and* $E\left[\varepsilon_1^{4+\delta_1}\right] < \infty$ *for some* $\delta_1 > 0$*. In addition,* $P\{\psi_t > 0\} = 1$ *for all* $t \ge 1$.

(iii) Suppose that $\{\psi_t\}$ *has a common marginal density* $f(\cdot)$ *and that* g_1, g_2 *and* f *have continuous derivatives of up to the second order and are bounded on the interior of* S_ω*. In addition,* $\inf_{\psi\in S_\omega} f(\psi) > 0$.

Assumption B.3. *(i) Suppose that Assumption A.1(ii) above holds and that the second derivative,* $k^{(2)}(u)$*, is continuous.*

(ii) Suppose that the nonnegative weight function $\omega(\cdot)$ *is continuous and bounded. In addition, the support* S_ω *is compact.*

Assumption B.4. *Recall* $z_t = x_{t-1} - g_2(\psi_{t-1})$ *and* $\psi_t = \gamma x_{t-1} + g(\psi_{t-1})$*. Let* $E\left(|\psi_t|^{4+\delta_2}\right) < \infty$ *and* $E\left(|z_t|^{4+\delta_2}\right) < \infty$ *for some* $\delta_2 > 0$.

Additionally, suppose that $\max_{t\ge 1} E\left[\left|\overline{\psi}_{t,0} - \psi_{t,0}\right|\right] < \overline{B} < \infty$ *and* $\sup_{x\in S_\omega} |g_j(x)| \le B_g < \infty$.

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9 (2), 385–426.

Bauwens, L., Giot, P., Grammig, J., Veredas, D., 2004. A comparison of financial duration models via density forecasts. International Journal of Forecasting 20 (4), 589–609.

Bauwens, L., Veredas, D., 2004. The stochastic conditional duration model: a latent variable model for the analysis of financial durations. Journal of Econometrics 119 (2), 381–412.

Chen, S., 2000. Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 52 (3), 471–480.

De Luca, G., Gallo, G., 2004. Mixture processes for financial intradaily durations. Studies in Nonlinear Dynamics and Econometrics 8 (2), 1–18.

Drost, F. C., Werker, B., 2004. Semiparametric duration models. Journal of Business & Economic Statistics 22 (1), 40–50.

Engle, R., Russell, J., 1997. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. Journal of Empirical Finance 4 (2-3), 187–212.

Engle, R., Russell, J., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162.

Fernandes, M., Grammig, J., 2005. Nonparametric specification tests for conditional duration models. Journal of Econometrics 127 (1), 35–68.

Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. Journal of Econometrics 130 (1), 1–23.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, London.

Gao, J., King, M. L., 2004. Adaptive testing in continuous-time diffusion models. Econometric Theory 20 (05), 844–882.

Giot, P., 2000. Time transformations, intraday data and volatility models. Journal of Computational Finance 4, 31–62.

Glaser, R., 1980. Bathtub and related failure rate characterizations. Journal of the American Statistical Association 75 (371), 667–672.

Grammig, J., Maurer, K., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. The Econometrics Journal 3 (1), 16–38.

Györfi, L., Härdle, W., Sarda, P., Vieu, P., 1989. Nonparametric Curve Estimation from Time Series. Vol. 60 of Lecture Notes in Statistics. Springer-Verlag, Berlin.

Horowitz, J., Spokoiny, V., 2001. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69 (3), 599–631.

Jones, M., 1993. Simple boundary correction for kernel density estimation. Statistics and Computing 3 (3), 135–146.

Li, Q., Racine, J., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press Princeton, NJ.

Meitz, M., Teräsvirta, T., 2006. Evaluating models of autoregressive conditional duration. Journal of Business & Economic Statistics 24 (1), 104–124.

Prigent, J., Renault, O., Scaillet, O., 2000. An autoregressive conditional binomial option pricing model. LSE Financial Markets Group Discussion Paper Series.

Silverman, B., 1986. Density estimation for statistics and data analysis. Chapman & Hall/CRC.

Tjøstheim, D., 1994. Nonlinear time series: a selective review. Scandinavian Journal of Statistics 21, 97–130.

Wongsaart, P., Gao, J., Allen, D., 2011. On the third generation autoregressive conditional duration model. Working paper available at https://sites.google.com/site/saartpw/research.

Zhang, M. Y., Russell, J. R., Tsay, R., 2001. A nonlinear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics 104 (1), 179–207.