ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

Econometric Time Series Specification Testing in a Class of Multiplicative Error Models

Patrick W. Saart, Jiti Gao and Nam Hyun Kim

January 2014

Working Paper 01/14

Econometric Time Series Specification Testing in a Class of Multiplicative Error Models

Patrick W. Saart
Department of Mathematics and Statistics
University of Canterbury, New Zealand

Jiti Gao[1]
Department of Econometrics and Business Statistics
Monash University, Australia

Nam Hyun Kim
Department of Economics
University of Konstanz, Germany

Abstract

In recent years, analysis of financial time series has focused largely on data related to market trading activity. Apart from modelling the conditional variance of returns within the GARCH family of models, presently attention has also been devoted to other market variables, especially volumes, number of trades and durations. The financial econometrics literature has focused on Multiplicative Error Models (MEMs), which are considered particularly suited for modelling certain financial variables. The paper establishes an econometric specification approach for MEMs. In the literature, several procedures are available to perform specification testing for MEMs, but the proposed specification testing method is particularly useful within the context of the MEMs of financial duration. The paper makes a number of important theoretical contributions. Both the proposed specification testing method and the associated theory are established and evaluated through simulations and real data examples.

JEL Classification: C14, C41, F31.

Keyword: Financial duration process; Nonnegative time series; Nonparametric kernel estimation; Semiparametric mixture model.

1. Introduction

In financial econometrics, the MEM is first introduced by Engle (2002) as a general class of time series models for positive-valued random variables, which are decomposed into the product of their conditional mean and a positive-valued error term. Alternatively, the MEM can be classified as an autoregressive conditional mean model where the conditional mean of a distribution is assumed to follow a stochastic process. The idea of the MEM is well-known in financial econometrics since it originates from the structure of the autoregressive conditional heteroskedasticity (ARCH) model introduced by Engle (1982) and the stochastic volatility model proposed by Taylor (1982), where the conditional variance is dynamically parameterized and multiplicatively interacts with an innovation term.

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

The MEM can be conveniently applied to any positive-valued continuous process.[2] For instance, to model the inter-temporal dynamics of trading volumes, Manganelli (2005) establishes a special case of the MEM called an autoregressive conditional volume process. Furthermore, Chou (2005) proposes an alternative specification known as the conditional autoregressive range model to study the dynamics of the high/low range of asset prices within fixed time intervals. Nonetheless, the most well-known application of the MEM by far is in the econometric analysis of high-frequency data. The MEM is first used within this context by Engle and Russell (1998) to model the clustering behavior of the waiting times between financial events, for instance market trading and changes in asset prices. Hereafter, let us refer to these waiting times as financial durations. The resulting model is referred to in the literature as the autoregressive conditional duration (ACD) model. The ACD class of models 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 two arrival times, i.e. $x_t = i_t - i_{t-1}$, measures the length of time commonly known as the duration, where $\{x_t\}$ 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 model assumes a multiplicative error model of x_t of the form

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

where $\psi_t \equiv \vartheta(x_{t-1}, \ldots, x_{t-p}, \psi_{t-1}, \ldots, \psi_{t-q})$ with $\vartheta : \mathbb{R}_+^p \times \mathbb{R}_+^q \to \mathbb{R}_+$ is a strictly positive–valued function and $\{\varepsilon_t\}$ is a strictly stationary time series with non-negative support density $f(\varepsilon; \phi)$, in which ϕ is a vector of the parameters and $E[\varepsilon_1] = 1$.

In practice, various procedures are available to perform hypothesis testing for the MEM. The Portmanteau test for residuals is often performed based on the classical Box-Pierce and Ljung-Box tests. Furthermore, there are several independence tests, for example the BDS tests proposed by Broock et al. (1996) and the omnibus test of Hong and Lee (2003) based on generalized spectral densities. For the MEM of conditional duration, some misspecification tests have also been proposed in the literature. Li and Yu (2003) derive a portmanteau test that can be used to evaluate the adequacy of an estimated ACD model. On the one hand, Meitz and Teräsvirta (2006) present a framework for evaluating models of conditional duration based on Lagrange multiplier misspecification tests of the functional form of the conditional mean duration, while, on the other hand, Hong and Lee (2011) do so based on the generalized spectral derivative approach. Moreover, a number of recent studies consider explicit tests on the distribution of the error terms. Diebold et al. (1998), for instance, propose a density evaluation method based on the Rosenblatt (1952)'s probability integral transform. Bauwens et al. (2004) apply this concept to compare a group of alternative financial duration models. A formal test against distributional misspecification is proposed by Fernandes and Grammig (2005) based on gauging the closeness between parametric and nonparametric estimates of the density function of the residuals as in Aït-Sahalia (1996).

[2] A point of view which is publicized by Engle (2002).

Within the context of the MEM, the misspecification in the distribution of the error term have a number of important implications. Despite its asymptotic consistency, Grammig and Maurer (2000) find that the quasi maximum likelihood (QML) method based on distributions belonging to the standard gamma family, for instance the exponential, fails to produce sound finite-sample results even in quite large samples. There are also other practical implications that are more directly related to the MEM for financial durations. The error term in the ACD class of models is usually interpreted as the standardized duration. Engle and Russell (1998) express the conditional intensity of financial events as an accelerated failure time model that depends very much on the baseline hazard, which is determined in turn by the distribution of the standardized duration. Ghysels et al. (2004) find that a misspecification of the baseline distribution had a serious implication for models that attempt to uncover the link between duration and volatility.[3] Furthermore, the success of option pricing and risk management procedures based on intraday volatility estimated from price duration models depend heavily on the appropriate specification of the baseline hazard rate function.[4]

Moreover, other interesting and important applications of the distribution test for the ACD class of MEM can be constructed based on the well-known time-change theorem of point processes. From the point of view of an intensity-based model of point processes, it has been shown that any point process whose compensator has continuous paths that increase to ∞ (e.g., the well-known Hawkes process) can be time-scaled to a standard Poisson process.[5] The scaling (also referred to as time-changing) is given by the inverse to the compensator. Thus, a large class of point processes with totally inaccessible event times can be viewed as the standard Poisson process with the appropriate time scale. Clearly, a corresponding time-scaling can be obtained from the duration-based point of view where exogenous influences (e.g. the intraday trading patterns) and the self-excited dynamics are modeled by the ACD framework. In this case, the rejection of the null hypothesis of the exponential distribution of the error term in an ACD model of transaction duration, for example, provides important evidence against the Poisson process of transaction arrivals. Such outcome has a few important implications. Firstly, the misspecification test developed here can be used in evaluating conditional mean and the conditional duration models.[6] Secondly, the rejection of the null hypothesis could be the evidence against the exogeneity of the trade arrival process, which is a common assumption in the financial literature.[7] Similar to the (intensity-based) Hawkes model, the ACD framework is able to model a deterministic function of time that accounts for the intensity of arrival of exogenous events and the endogenous feedback mechanism, i.e. the memory of the process. However, other exogenous influences, which are often modelled as the doubly-stochastic processes, are ignored.

---

[3]See also Engle (2000).

[4]See Giot (2000), for example.

[5]See Giesecke and Tomecek (2005), for example.

[6]A similar use of the time-change theorem can be found in Bowsher (2007) and Giesecke and Kim (2011), for example.

[7]See Dufour and Engle (2000), for example, and the references therein.

Having outlined its statistical and empirical relevance, the main objective of this paper is to establish an alternative nonparametric testing procedure to specify the probability distribution of the error term involved in the MEM. The procedure will be established based on an intricate network of recent developments in non– and semi–parametric time series. This can be outlined as the following:

(1) In order to derive the test statistic, we employ a smoothed version of the parametric alternative that is local-linearly weighted to avoid undersmoothing requirement which is encountered by Fernandes and Grammig (2005) and Aït-Sahalia (1996). (2) A number of existing studies have shown that bandwidth selection in nonparametric kernel testing is not a straightforward matter (see, for example, Gao (2007), and Gao and Gijbels (2008)). Generally, one can distinguish two approaches which deal with the choice of bandwidth parameter in nonparametric and semiparametric kernel methods for testing hypotheses. The first approach is to use an estimation-based optimal bandwidth value from, for example, a cross-validation. However, this may lead to poor performance in finite sample studies since the estimation-based optimal bandwidth may not necessarily guarantee the optimality of the test. The second approach is to select the bandwidth among a set of pre-specified values. The research in this paper entails a methodological contribution to the MEM literature by applying Horowitz and Spokoiny (2001) approach, which is to simultaneously consider a family of test statistics associated with a set of possible smoothing parameters values. The proposed test rejects the null hypothesis if at least one of the test statistics is sufficiently large. (3) For applications of the MEM in financial time series, the mean residuals may have substantial mass close to the zero boundary. In the current paper, an alternative version of the test statistic is developed such that the boundary problem is considered through a method of boundary correction as discussed in Jones (1993). (4) In recent years, nonparametric and semiparametric estimation with generated covariate(s) has become an important research venue in econometrics and statistics.[8] Nonetheless, the research in this paper makes a headway in statistical inferences (hypothesis testing in particular) with nonparametrically and semiparametrically generated variables. An important feature about the nonparametric specification testing of the MEM introduced in this paper is the fact that it is implemented based on an estimate, namely the MEM residual, to test the probability distribution of the true but unobservable error process.

The article is organized as follows. In Section 2 we present the testing procedure and a number of important asymptotic results. In Section 3 we discuss and experimentally illustrate the applicability of the test and the associated asymptotic results to the multiplicative-error modeling of financial duration. In Section 4 we consider a special version of the test for cases by which the error terms have substantial mass close to the zero-boundary. In Section 5 we illustrate the use the tests using data from the New York Stock Exchange (NYSE) and the Australian Stock Exchange (ASX). Finally, we conclude in Section 6 and provide proof in an appendix.

[8] See Sperlich (2009) and Mammen et al. (2012), for example.

2. Testing the Marginal Density Function

In the current section, we first derive the test statistic and its asymptotic distribution. We then propose an adaptive test procedure and discuss the consistency of the test.

2.1. Test Statistic and Asymptotic Distribution

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

$$\mathcal{H}_0 : f(\varepsilon) = f(\varepsilon, \theta_0) \tag{2.1}$$

against a sequence of local alternatives

$$\mathcal{H}_1 : f(\varepsilon) = f(\varepsilon, \theta_1) + C_T \Delta_T(\varepsilon), \quad \theta_1 \in \Theta, \tag{2.2}$$

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 such a way that the alternative function is still a probability density under $\mathcal{H}_1$. Specifically, we consider the case where

$$\Delta_T(\varepsilon) = \varphi(\varepsilon) - f(\varepsilon, \theta_1) \tag{2.3}$$

so that the alternative hypothesis in (2.2) is a semiparametric mixture density

$$\mathcal{H}_1 : f(\varepsilon) = (1 - C_T) f(\varepsilon, \theta_1) + C_T \varphi(\varepsilon), \tag{2.4}$$

where φ denotes an unknown density function. The alternative hypothesis in (2.4) includes as special cases

$$\mathcal{H}_1' : f(\varepsilon) = f(\varepsilon, \theta_1) \text{ and } \mathcal{H}_1'' : f(\varepsilon) = (1 - C) f(\varepsilon, \theta_1) + C\varphi(\varepsilon), \tag{2.5}$$

which are obtained when $C_T = 0$ and $C_T = C$ for $0 < C < 1$, respectively.

To discuss our test statistic, let us first define the distance function $D(f, \theta) = \int (f(\varepsilon) - f(\varepsilon, \theta))^2 f(\varepsilon) d\varepsilon$. The main idea behind the test statistic considered in this paper is to compare a consistent nonparametric density estimator directly with a parametric density in question. Let $\widehat{\varepsilon}_t$ denote an estimate of ε_t, that satisfies Assumption 2.1 below. An alternative approach to estimating $D(f, \theta)$ is to do so based on the following quantity:

$$D(\widehat{f}, \widetilde{\theta}) = \int (\widehat{f}(\widehat{\varepsilon}) - f(\widehat{\varepsilon}, \widetilde{\theta}))^2 \widehat{f}(\widehat{\varepsilon}) \, d\widehat{\varepsilon}, \tag{2.6}$$

where $\widehat{\varepsilon}$ denotes an estimated version of ε,

$$\widehat{f}(\widehat{\varepsilon}) = \frac{1}{T} \sum_{t=1}^{T} k_h(\widehat{\varepsilon} - \widehat{\varepsilon}_t) \tag{2.7}$$

denotes the standard kernel density estimator with $k_h(\cdot) = h^{-1}k(\cdot/h)$, $k(\cdot)$, h is a bandwidth parameter and $\widetilde{\theta}$ is a consistent estimator of θ. Nonetheless, obtaining an asymptotically normal distribution with zero mean for $D(\widehat{f}, \widetilde{\theta})$ in this case requires $\lim_{T\to\infty} Th^{4.5} = 0$, which implies undersmoothing.[9] An alternative to avoid the undersmoothing issue is to write

$$\widetilde{D}(\widehat{f}, \widetilde{\theta}) = \int (\widehat{f}(\widehat{\varepsilon}) - \widetilde{f}(\widehat{\varepsilon}, \widetilde{\theta}))^2 \widehat{f}(\widehat{\varepsilon})\, d\widehat{\varepsilon}, \tag{2.8}$$

where

$$\widetilde{f}(\widehat{\varepsilon}, \widetilde{\theta}) = \sum_{t=1}^{T} w_t(\widehat{\varepsilon}) f(\widehat{\varepsilon}_t, \widetilde{\theta}) \tag{2.9}$$

denotes a nonparametrically smoothed version of $f(\widehat{\varepsilon}, \theta)$ with

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

and $s_r(\widehat{\varepsilon}) = (1/T) \sum_{s=1}^{T} k_h(\widehat{\varepsilon} - \widehat{\varepsilon}_s)(\widehat{\varepsilon} - \widehat{\varepsilon}_s)^r$ for $r = 0, 1, 2$. Gao and King (2004) show that the use of the difference $\widehat{f}(\widehat{\varepsilon}) - \widetilde{f}(\widehat{\varepsilon}, \widetilde{\theta})$ can avoid undersmoothing in the sense that we can still assume $\limsup_{T\to\infty} Th^5 < \infty$. This suggests introducing a test statistic of the form

$$\breve{N}_{T,\widehat{\varepsilon}} \equiv \breve{N}_{T,\widehat{\varepsilon}}(h) = Th \int (\widehat{f}(\widehat{\varepsilon}) - \widetilde{f}(\widehat{\varepsilon}, \widetilde{\theta}))^2 \; \widehat{f}(\widehat{\varepsilon})\, d\widehat{\varepsilon}. \tag{2.10}$$

In the current paper, we will take (2.10) as a starting point and derive our test statistic based on the following discretized version

$$\widehat{N}_{T,\widehat{\varepsilon}}(h) = \frac{Th}{T} \sum_{t=1}^{T} \left(\widehat{f}(\widehat{\varepsilon}_t) - \widetilde{f}(\widehat{\varepsilon}_t, \widetilde{\theta}) \right)^2. \tag{2.11}$$

The main difficulty involved in the establishment of an asymptotic theory for $\widehat{N}_{T,\widehat{\varepsilon}}(h)$ is the fact that it is implemented based on the residual in order to test the probability distribution of the true but unobservable error term. In the remainder of this section, we present key asymptotic results that help us overcome this difficulty in the mathematical proof of Theorems 2.1 to 2.4 below. Let us begin with the following notation. Hereafter, let $\widehat{N}_{T,\varepsilon}(h) = h \sum_{t=1}^{T} \left(\widehat{f}(\varepsilon_t) - \widetilde{f}(\varepsilon_t, \widetilde{\theta}) \right)^2$ and

$$L_{T,\varepsilon}(h) = \frac{\widehat{N}_{T,\varepsilon}(h) - \mu_0}{\sqrt{h}\sigma_0}, \tag{2.12}$$

where $\mu_0 = R(k) \int_{\infty}^{\infty} f^2(u)du$ and $\sigma_0^2 = 2k^{(4)}(0) \int f^4(u)du$, in which $R(k) = k^2(u)du$, and let

$$\widehat{L}_{T,\varepsilon}(h) = \frac{\widehat{N}_{T,\varepsilon}(h) - \widehat{\mu}_{T,\varepsilon}(h)}{\sqrt{\widehat{h}}\widehat{\sigma}_{T,\varepsilon}(h)}, \tag{2.13}$$

[9] See Fan (1994), and Fernandes and Grammig (2005), for example.

where $\widehat{\mu}_{T,\varepsilon}$ and $\widehat{\sigma}^2_{T,\varepsilon}$ are obtained by replacing $\widehat{\varepsilon}$ with ε in (2.16).

The key technical assumption required for the derivation of Lemma 2.1 is Assumption 2.1 below, which is a high-level restriction on the accuracy of the estimator $\widehat{\varepsilon}$. Assumption 2.1 requires $\widehat{\varepsilon}$ to be uniformly consistent with a rate of convergence that is at least as fast as that of the bandwidth parameter converging to zero. meanwhile, we also introduce some other conditions in Assumptions 2.2 and 2.3 below.

Assumption 2.1. *Assume that there exists an estimator $\widehat{\varepsilon}_t$ of ε_t such that*

$$\sup_{t\geq 1} |\widehat{\varepsilon}_t - \varepsilon_t| = o_P(T^{-\delta}) \tag{2.14}$$

for some δ, where $\delta > 0$ is chosen such that $T^{\delta}h \to \infty$.

Assumption 2.2. *(i) Suppose that 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 $r = 1, 2$.*

(ii) Let $k_2 = \int_{-\infty}^{\infty} u^2 k(u)du > 0$. Suppose that k has an absolutely integrable Fourier transform with $\int_{-\infty}^{\infty} u^2 k\left(u\right) du < \infty$ and $\int_{-\infty}^{\infty} k^2(u)du < \infty$.

(iii) Let $H(\cdot)$ denote a bounded function. Suppose that $k(u) \leq H(u)$ and $\left|k^{(q)}(u)\right| \leq H(u)$ for $q = 1, 2$.

(iv) The error process $\{\varepsilon_t\}$ is strictly stationary and α–mixing with mixing coefficient $\alpha_\varepsilon(k)$ satisfying $\alpha_\varepsilon(k) \leq C_\varepsilon q_\varepsilon^k$, where $0 < C_\varepsilon < \infty$ and $0 < q_\varepsilon < 1$.

(v) Let $\{\varepsilon_t\}$ satisfy $E\left[\varepsilon_1\right] = 1$ and $E\left[\varepsilon_1^{4+\delta_1}\right] < \infty$ for some $\delta_1 > 0$.

(vi) Assume that the first three derivatives of $f(x)$ are continuous in $x \in R^1 = (-\infty, \infty)$ and that $f(x) > c_f > 0$ for $x \in D$, any compact subset of R^1 for some $c_f > 0$.

(vii) Assume that $E\left[\left|f^{(r)}(\varepsilon_1)\right|\right] < \infty$ for $1 \leq r \leq 3$. In addition, the initial random variable ε_0 is distributed from $f(\varepsilon)$ with $0 < \int f^2(u)du < \infty$ and $0 < \int f^4(u)du < \infty$.

Assumption 2.3. *(i) The parameter space $\Theta \subset R^q$ is compact.*

(ii) In a neighborhood of the true parameter θ_0, $f(\varepsilon, \theta)$ is twice continuously differentiable in θ; $E[(\partial f(\varepsilon_1, \theta)/\partial\theta)(\partial f(\varepsilon_1, \theta)/\partial\theta)^\tau]$ is full rank. In addition, suppose that $G(\varepsilon)$ is a positive and integrable function with $E[G(\varepsilon_1)] < \infty$ such that $\sup_{\theta\in\Theta} |f(\varepsilon_1, \theta)|^2 \leq G(\varepsilon_1)$ and $\sup_{\theta\in\Theta} ||\nabla_\theta^j f(\varepsilon_1, \theta)||^2 \leq G(\varepsilon_1)$ with probability one for $j = 1, 2, 3$, where $||B||^2 = \sum_{i=1}^q \sum_{j=1}^q b_{ij}^2$ for $B = \{b_{ij}\}_{1\leq i,j\leq q}$.

(iii) 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.

(a) For each $\theta \in \Theta$, $f(x, \theta)$ is continuous with respect to $x \in R^1$.

(b) 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).

(iv) 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.

(v) 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.

(vi) Suppose that $\Delta_T(x)$ is continuous in x and satisfies $\int_{-\infty}^{\infty} \Delta_T(x)dx = 0$ for all $T \geq 1$.

With the exception of Assumption 2.2(iv), Assumption 2.2 is a set of standard regularity (and smoothness) conditions for kernel-type nonparametric regression. Assumption 2.2(iv) assumes the α-mixing condition, which is weaker than the β-mixing condition. Assumptions 2.2(i) and 2.2(ii) are quite general in that they allow for the use of the standard normal kernel. The conditions in Assumption 2.3 are consistent with assumptions A0, A1 and A3 of Aït-Sahalia (1996), which are also required by Fernandes and Grammig (2005).

The proofs of Theorems 2.1–2.4 are based on an application of Lemma 2.1 below.

Lemma 2.1. *Let Assumptions 2.1 to 2.2 below hold. Then, we have, as $T \to \infty$,*

$$\widehat{L}_{T,\widehat{\varepsilon}}(h) = \widehat{L}_{T,\varepsilon}(h) + o_P(T^{-\delta})$$

uniformly in $h \in H_T$ for some δ, where $\delta > 0$ is chosen such that $T^{\delta} h \to \infty$.

The first result of this section establishes an asymptotic distribution for the test statistic under $\mathcal{H}_0$.

Theorem 2.1. *Let Assumptions A.1 to A.3 listed in Appendix A below hold. Suppose that there is a random data-driven $\widehat{h}$ such that $\frac{\widehat{h}}{h} - 1 \to_P 0$ as $T \to \infty$. Then, under $\mathcal{H}_0$ we have*

$$\widehat{L}_{T,\widehat{\varepsilon}}(\widehat{h}) = \frac{\widehat{N}_{T,\widehat{\varepsilon}}(\widehat{h}) - \widehat{\mu}_{T,\widehat{\varepsilon}}(\widehat{h})}{\sqrt{\widehat{h}}\widehat{\sigma}_{T,\widehat{\varepsilon}}(\widehat{h})} \underset{D}{\to} N(0,1) \text{ as } T \to \infty, \tag{2.15}$$

where

$$\widehat{\mu}_{T,\widehat{\varepsilon}}(h) = R(k) \cdot \left(\frac{1}{T} \sum_{i=1}^{T} \widehat{f}(\widehat{\varepsilon}_t) \right) \text{ and } \widehat{\sigma}^2_{T,\widehat{\varepsilon}}(h) = 2k^{(4)}(0) \cdot \left(\frac{1}{T} \sum_{t=1}^{T} \widehat{f}^3(\widehat{\varepsilon}_t) \right), \tag{2.16}$$

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 $k^{(4)}(0) = \int_{-\infty}^{\infty} L^2(u)\, du$ with $L(u) = \int_{-\infty}^{\infty} k(y)k(u+y)\, dy$.

2.2. An Adaptive Test Procedure

A common approach for the choice of a bandwidth parameter in the kernel method for hypothesis testing is to use an optimal bandwidth value. For example, in order to implement their nonparametric specification tests, Fernandes and Grammig (2005) propose using a discretized version of $D(\widehat{f}, \tilde{\theta})$ together with an (undersmoothing-adjusted) theoretically optimal bandwidth. However, this may lead to a poor performance in finite sample studies since the estimation-based optimal bandwidth may not necessarily guarantee the optimality of the test.[10] To address this problem, the current paper follows the approach introduced by

[10] Gao and Gijbels (2008).

Horowitz and Spokoiny (2001) for testing nonparametric regression functions. In particular, we suggest using

$$L_{\widehat{\varepsilon}}^* \equiv \max_{h \in H_T} \widehat{L}_{T,\widehat{\varepsilon}}(h) = \max_{h \in H_T} \frac{\widehat{N}_{T,\widehat{\varepsilon}}(h) - \widehat{\mu}_{T,\widehat{\varepsilon}}(h)}{\sqrt{h}\widehat{\sigma}_{T,\widehat{\varepsilon}}(h)}, \tag{2.17}$$

where $\widehat{N}_{T,\widehat{\varepsilon}}$, $\widehat{\mu}_{T,\widehat{\varepsilon}}$, and $\widehat{\sigma}_{T,\widehat{\varepsilon}}$ are as specified in Theorem 2.1, and

$$H_T = \{h = h_{\max} a^k : h \geq h_{\min},\ k = 0, 1, 2, \ldots\}, \tag{2.18}$$

in which $0 < h_{\min} < h_{\max}$, $0 < a < 1$ and $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$. A more detailed discussion on the conditions on $h_{\max}$ and $h_{\min}$ can be found in Section A.1.

Below, let us introduce the method for 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 and data generated from a known marginal density.

Step 2.1: Given a non-negative time series process, perform an estimation of the MEM to obtain $\widehat{\varepsilon}_t$.

Step 2.2: Estimate the true value θ_0 based on $\widehat{\varepsilon}_t$. Denote the resulting estimate by $\widehat{\theta}$.

Step 2.3: Compute $L_{\widehat{\varepsilon}}^*$ based on $\widehat{\varepsilon}_t$ and $\widehat{\theta}$.

Step 2.4: Repeat Steps 2.1 to 2.3 Q number of times to obtain $L_{\widehat{\varepsilon},q}^*$ for $q = 1, 2, \ldots, Q$. Compute the simulated critical value l_α^* as the $(1-\alpha)$ percentile of the Q values of $L_{\widehat{\varepsilon}}^*$.

Theorem 2.2 below is essential in order to ensure the statistical validity of the simulation scheme discussed above. The main result on the behavior of the test statistic $L_{\widehat{\varepsilon}}^*$ under $\mathcal{H}_0$ is that l_α^* is an asymptotically correct α-level critical value under any model in $\mathcal{H}_0$.

Theorem 2.2. *Let Assumptions A.1 to A.3 listed in Appendix A below hold. Then, under $\mathcal{H}_0$, we have*

$$\lim_{T \to \infty} P(L_{\widehat{\varepsilon}}^* > l_\alpha^*) = \alpha. \tag{2.19}$$

Below, let us shift our focus to the consistency of $L_{\widehat{\varepsilon}}^*$ against the alternatives discussed previously. We will concentrate first on the fixed alternatives. Let $\mathcal{F} = \{f(\cdot, \theta) : \theta \in \Theta\}$ be a set of density functions that satisfy Assumption 2.2(iv) below, and let $F(\theta) = (f(\varepsilon_1, \theta), \ldots, f(\varepsilon_T, \theta))^\tau$ and $\bar{f} = (f(\varepsilon_1), \ldots, f(\varepsilon_T))^\tau$. The distance between f and $\mathcal{F}$ can then 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 2.3. *Let Assumptions A.1 to A.3 listed in Appendix A below 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 we have*

$$\lim_{T\to\infty} P(L^*_{\widehat{\varepsilon}} > l^*_\alpha) = 1. \tag{2.20}$$

Theorem 2.3 shows that if $\mathcal{H}_0$ is false, then $\rho(f,\mathcal{F}) \geq \mathcal{C}_\rho$ for all sufficiently large values of T and some $C_\rho > 0$. A consistent test will reject a false $\mathcal{H}_0$ with probability approaching one as $T \to \infty$.

In addition to studying the fixed alternative in Theorem 2.3, we show that $L^*_{\widehat{\varepsilon}}$ is consistent against a sequence of local alternatives of the following form:

$$f_T(\varepsilon) = f(\varepsilon, \theta_1) + C_T \Delta_T(\varepsilon). \tag{2.21}$$

The 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 as:

$$\frac{1}{T}\|\bar{f}_T - F(\theta_1)\|^2 = \frac{C_T^2}{T}\|\bar{\Delta}_T\|^2 = C_T^2 \left(\frac{1}{T}\sum_{t=1}^{T} |\Delta_T(\varepsilon_t)|^2\right). \tag{2.22}$$

To ensure that the rate of convergence of $\bar{f}_T$ 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 $0 < \delta < \infty$, $\Delta_T(\varepsilon)$ is a continuous function which is normalized so that we have

$$\lim_{T\to\infty} P\left(\frac{1}{T}\sum_{t=1}^{T} |\Delta_T(\varepsilon_t)|^2 \geq \delta\right) = 1. \tag{2.23}$$

Theorem 2.4. *Let Assumptions A.1 to A.3 listed in Appendix A below hold. In addition, let f_T satisfy (2.21) with $C_T \geq CT^{-1/2}\sqrt{\log\log T}$ for some constant $C > 0$ and let condition (2.23) hold. Then, we have*

$$\lim_{T\to\infty} P(L^*_{\widehat{\varepsilon}} > l^*_\alpha) = 1. \tag{2.24}$$

The establishments and the proofs of Theorems 2.1–2.4 are based on the high–level assumption in Assumption 2.1. In Section 3 below, Assumption 2.1 is naturally satisfied when the multiplicative error models reduce to two classes of autoregressive conditional duration models. Theorems 2.2–2.4 become Corollaries 3.1–3.3 below.

3. Hypothesis Testing for the MEM of Financial Duration

In this section, we consider two classes of autoregressive conditional duration models, and then illustrate the practical applicability of our asymptotic theory and testing procedure within the context of the MEMs of the financial durations.

3.1. Main Asymptotic Results

The multiplicative error model in (1.1) leads directly to

$$\varepsilon_t = \frac{x_t}{\psi_t}, \tag{3.1}$$

which is often referred to in the ACD literature as the standardized duration. The flexibility of the ACD class of models depends partially on the design of the conditional duration, which is determined in this paper by the functional form of ϑ. For instance, the ACD model of Engle and Russell (1998) (hereafter ER-ACD) assumes a specification of the form

$$\psi_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{j=1}^{q} \beta_j \psi_{t-j}. \tag{3.2}$$

The simplest specification of the conditional duration suggested for the ER-ACD class of models is the ER-ACD(1,1) by which

$$\psi_t = \alpha x_{t-1} + \beta \psi_{t-1}. \tag{3.3}$$

Furthermore, the simplest form of the semiparametric ACD model proposed by Saart et al. (2013), i.e. the so-called SEMI-ACD(1,1) model, defines

$$\psi_t = \gamma x_{t-1} + g\left(\psi_{t-1}\right), \tag{3.4}$$

where $g(\cdot)$ satisfies the Lipschitz-type condition stated in Assumption B.1 in Appendix B.2. In order to obtain some additional flexibility, a few other parametric functional forms of the conditional mean have also been considered in the literature.[11] Although it can be shown that the testing procedure developed in Section 2 is applicable to most of these models, for the sake of illustration, in the current paper we concentrate only on the SEMI-ACD and the ER-ACD models. It should be noted, however, that the SEMI-ACD model ensures flexibility not only in the error distribution, but also in the conditional duration and should, therefore, be the most suitable in minimizing the impact of a potential joint hypothesis problem in multiplicative error models.

By letting $\widehat{\psi}_t$ denote the estimate of the conditional duration either from the SEMI-ACD and from the ER-ACD models (below we will respectively denote these as $\widehat{\psi}_{t,m^*}$ and $\widehat{\psi}_{t,Qml}$), expression (3.1) suggests that[12]

$$\left|\widehat{\varepsilon}_t - \varepsilon_t\right| = \left|\frac{x_t}{\widehat{\psi}_t} - \frac{x_t}{\psi_t}\right| \;=\; x_t \left|\widehat{\psi}_t - \psi_t\right| \left\{\frac{\psi_t}{\widehat{\psi}_t}\right\}. \tag{3.5}$$

This, along with the boundedness of x, suggests that the uniform consistency of $\widehat{\psi}$ should immediately lead to a similar mode of convergence of $\widehat{\varepsilon}$. Therefore, in order to establish asymptotic properties for the adaptive test procedure within the ACD framework, we first derive the following uniform convergence results.

[11] See Pacurar (2008) for an excellent review.

[12] See Section B.1 for a detailed review of the the estimation algorithm of the SEMI-ACD model introduced in Saart et al. (2013).

Lemma 3.1. *Uniform Convergence of the SEMI-ACD Model: Let Assumptions B.1 to B.3 hold. Then we have*

$$\left|\widehat{\psi}_{t,m^*} - \psi_t\right| \leq \sup_{h\in H_T} \max_{t\geq m+1} \left|\widehat{\psi}_{t,m^*} - \psi_t\right| = o_P\left\{T^{-\frac{1}{4}}\right\}, \tag{3.6}$$

where $m^* = \left[C_G \cdot \log\left(T^{\frac{1}{4}}\right)\right]$ *for some* $C_G = \frac{1}{\log(G^{-1})}$ *and* $[x] \leq x$ *denotes the largest integer part of* x*. In addition, we have*

$$\sup_{h\in H_T} \max_{1\leq t\leq T} |\widehat{\varepsilon}_{t,m^*} - \varepsilon_t| = o_P\left\{T^{-1/4}\right\}. \tag{3.7}$$

Lemma 3.2. *Uniform Convergence of the ER-ACD Model: Under the conditions of the ER-ACD model as introduced in Engle and Russell (1998), we have*

$$\left|\widehat{\psi}_{t,Qml} - \psi_t\right| \leq \max_{1<t\leq T} \left|\widehat{\psi}_{t,Qml} - \psi_t\right| = o_P\left\{T^{-\frac{1}{2}}\right\}. \tag{3.8}$$

In addition, we have

$$|\widehat{\varepsilon}_{t,Qml} - \varepsilon_t| \leq \max_{1<t\leq T} |\widehat{\varepsilon}_{t,Qml} - \varepsilon_t| = o_P\left\{T^{-\frac{1}{2}}\right\}. \tag{3.9}$$

Let $\{\varepsilon_t\}$ be the standardized duration process of a financial event. In addition, let $f_\varepsilon(\cdot)$ and $f_\varepsilon(\cdot,\theta)$ be the nonparametric and 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 will consider a testing procedure to test the null hypothesis

$$\mathcal{H}_0 : f_\varepsilon(\varepsilon) = f_\varepsilon(\varepsilon, \theta_0) \tag{3.10}$$

against a sequence of local alternatives, namely:

$$\mathcal{H}_1 : f_\varepsilon(\varepsilon) = f_\varepsilon(\varepsilon, \theta_1) + C_T \Delta_T(\varepsilon), \tag{3.11}$$

where $\theta_1 \in \Theta$, $0 \leq C_T \leq 1$, $\lim_{T\to\infty} C_T = 0$ and $\Delta_T(\varepsilon) = \varphi(\varepsilon) - f_\varepsilon(\varepsilon,\theta_1)$, so that

$$\mathcal{H}_1 : f_\varepsilon(\varepsilon) = (1 - C_T) f_\varepsilon(\varepsilon, \theta_1) + C_T \varphi(\varepsilon). \tag{3.12}$$

A couple of special cases, as seen previously, are the fixed alternatives of the forms

$$\mathcal{H}_1' : f_\varepsilon(\varepsilon) = f_\varepsilon(\varepsilon, \theta_1) \ \text{ and } \ \mathcal{H}_1'' : f_\varepsilon(\varepsilon) = (1 - C) f_\varepsilon(\varepsilon, \theta_1) + C\varphi(\varepsilon), \tag{3.13}$$

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

Since Lemmas 3.1 and 3.2 imply that Assumption 2.1 is satisfied in each of the cases, we are able to derive the corresponding asymptotic results to those of Theorems 2.2 to 2.4 that are specifically applicable to hypothesis testing for the MEM of financial duration as stated in the Corollaries below.

Corollary 3.1. *Let Assumptions B.1 to B.3 hold. Then, under $\mathcal{H}_0$, we have*

$$\lim_{T\to\infty} P(L^*_{\hat{\varepsilon}} > l^*_\alpha) = \alpha. \tag{3.14}$$

Corollary 3.2. *Let Assumptions B.1 to B.3 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 we have*

$$\lim_{T\to\infty} P(L^*_{\hat{\varepsilon}} > l^*_\alpha) = 1. \tag{3.15}$$

Corollary 3.3. *Let Assumptions B.1 to B.3 hold. In addition, let f_T satisfy (2.21) with $C_T \geq CT^{-1/2}\sqrt{\log\log T}$ for some constant $C > 0$ and let condition (2.23) hold. Then*

$$\lim_{T\to\infty} P(L^*_{\hat{\varepsilon}} > l^*_\alpha) = 1. \tag{3.16}$$

The proofs of these Corollaries are given in Section B.4.

3.2. Finite Sample Properties: Monte Carlo Studies

The main objective of the Monte Carlo exercises conducted in this section is to assess the performance of the above-discussed adaptive test procedure in finite sample. Note that hypothesis testing within the ACD framework is used here as an illustrative example. This is due especially to the fact that the uniform convergence in Lemmas 3.1 and 3.2 are readily derived. This makes the ACD framework an ideal venue for evaluating whether the procedure is able to perform the job even with a generated variable. While specific details about each exercise are presented in a relevant section, right below let us discuss some general issues.

To facilitate the study in this section, we simulate a SEMI-ACD model with the following dynamic structure of the duration process

$$\psi_t = \gamma x_{t-1} + \lambda \left(\frac{\psi_{t-1}}{1 + \psi^2_{t-1}} \right), \tag{3.17}$$

where $\gamma = 0.5$ and $\lambda = 0.75$. Hereafter, let us refer to the model in (3.17) as the Mackey-Glass ACD (MG-ACD) model. The functional form of g given above ensures that the simulated duration process $\{x_t\}$ is strictly stationary (see Section 2.4 of Tjøstheim (1994) for more detail). Furthermore, Lemma 3.4.4 and Theorem 3.4.10 of Györfi et al. (1989) can be used to show 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 that is essentially the key assumption behind most of the ACD class of models. [13]

Moreover, to examine the robustness of the test, we formulate models with data generating processes that exhibit hazard rate functions of various different shapes. To achieve this objective, we employ a family of three-parameter generalized gamma densities given by

$$f_{GG}(\varepsilon; \alpha, \beta, \delta) = \left\{ \frac{\delta \varepsilon^{\delta\alpha - 1}}{\beta^{\delta\alpha}\Gamma(\alpha)} \right\} \exp\left\{ -\left(\frac{\varepsilon}{\beta}\right)^\delta \right\} \quad \text{for} \quad \varepsilon \geq 0, \tag{3.18}$$

[13] It is true that the same applies for most ACD-type models in the literature (at least for β-mixing; Meitz and Teräsvirta (2006)) and this makes them equivalently good candidates.

where $\alpha > 0$, $\beta > 0$ and $\delta > 0$. Note that f_{GG} covers the following probability density functions of the gamma ($\delta = 1$), the exponential ($\alpha = 1$ and $\delta = 1$) and the Weibull distributions ($\alpha = 1$). To this end, Glaser (1980) derives a well–known result, which draws a connection between the sign of $\eta' = \frac{-f'(\varepsilon)}{f(\varepsilon)}$ and the shape of hazard rate function. In Table 3.1, we present some cases that will be useful for our study.

Table 3.1. 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)

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

All computations in this section are done in R. To compute the nonparametric estimates involved, we choose the normal kernel function $k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. In order to perform the hypothesis testing, we define H_T as in (2.18) with $h_{\min} = T^{-\frac{11}{36}}, h_{\max} = 2(\log\log T)^{-1}$ and $a = \frac{35}{36}$. The corresponding simulated critical values at the α-level are computed using the simulation scheme proposed in Section 2.2 with $Q = 1000$. The number of simulations used is 200. Three different sample sizes, namely $T - 1 \equiv N = 175, 300, 500$, are considered. [14]

3.2.1. Size Values of the Test

In this section, we consider the following null hypotheses:

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

[14] To be consistent with the sample sizes that are often encountered in financial studies, we have also considered larger numbers of observations, e.g. $T \geq 1000$. However, some of the results are not meaningful since these large numbers of observations often result in the power of the test of 1.

(b) $\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0)$ with $\theta_0 = (\alpha_0 = 2, \beta_0 = 1)$, implying 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 3.2. 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 3.3. The size of the test $L_{\widehat{\varepsilon}}^*$.

	$\mathcal{H}_0 : f(\varepsilon) = f_{GG}(\varepsilon, \theta_0)$		$\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0)$		$\mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \lambda_0)$	
T	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
176	0.065	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

Table 3.4. Distance $D(f_{\mathcal{H}_1}(\varepsilon))$.

Testing Examples	Denotation	Densities	α	δ	β	$D(f_{\mathcal{H}_1}(\varepsilon))$
Set 1	E	$f_E(\varepsilon, \lambda_0)$	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	$D(f_{\mathcal{H}_1}(\varepsilon))$
Set 2	E	$f_E(\varepsilon, \theta_0)$	·	·
	SD_{C1}	$(1 - C_1)f_E(\varepsilon, \lambda_1) + C_1\ \varphi(\varepsilon)$	0.15	0.0001
	SD_{C2}	$(1 - C_2)f_E(\varepsilon, \lambda_1) + C_2\ \varphi(\varepsilon)$	0.17	0.0003
	SD_{C3}	$(1 - C_3)f_E(\varepsilon, \lambda_1) + C_3\ \varphi(\varepsilon)$	0.19	0.0004

Tables 3.2 and 3.3 present the simulation results of the size values of L_ε^* and $L_{\widehat{\varepsilon}}^*$, respectively. In all cases, the simulated sizes obtained are quite close to their corresponding critical levels. In general, all the rates in Table 3.3 are similar to those of their ε-based counterparts in Table 3.2.

3.2.2. Power Values of the Test against a Fixed Alternative

In this section, we examine the power of the test for the following hypotheses

$$\text{Set 1} \quad \mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \lambda_0) \quad vs. \quad \mathcal{H}_1 : f(\varepsilon) = f_G(\varepsilon, \theta_j) \tag{3.19}$$

$$\text{Set 2} \quad \mathcal{H}_0 : f(\varepsilon) = f_E(\varepsilon, \lambda_0) \quad vs. \quad \mathcal{H}_1 : f(\varepsilon) = (1 - C_s)f_E(\varepsilon, \lambda_1) + C_s\ \varphi(\varepsilon), \tag{3.20}$$

where $j = 1, 2, 3$ and $s = 1, 2, 3$, $0 < C_s < 1$ and θ_j are vectors of the gamma parameters and $\varphi(\varepsilon)$ represents an unknown density function. The first set is constructed so that more 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 ones. Specifically, we look at testing the null hypothesis of $f_E(\varepsilon, \lambda_0)$ with a constant hazard rate function against the alternatives $f_G(\varepsilon, \theta_j)$ for $j = 1, 2, 3$ of which the hazard rate functions increase incrementally. Regarding the second set, in order to ensure that the alternative function in this case is still a probability density function and that it is fairly close to the exponential law under $\mathcal{H}_0$, we let

$$\varphi(\varepsilon) = \left\{ \frac{1}{\left(\frac{1}{4}\right)^{\frac{1}{4}}} \right\} \frac{\varepsilon^{-\frac{3}{4}}}{\Gamma\left(\frac{1}{4}\right)} \exp\left\{ -\frac{4}{\varepsilon} \right\}, \tag{3.21}$$

which is a gamma density with $\alpha = \beta = \frac{1}{4}$. Theoretically, the power of the test depends on the distance between $\mathcal{H}_0$ and $\mathcal{H}_1$. By letting $f_{\mathcal{H}_1}(\varepsilon)$ denote the density function under $\mathcal{H}_1$, we quantify this distance by

$$D(f_{\mathcal{H}_1}(\varepsilon)) = \int (f_E(\varepsilon, \lambda_0) - f_{\mathcal{H}_1}(\varepsilon))^2 f_E(\varepsilon, \lambda_0) d\varepsilon. \tag{3.22}$$

Table 3.4 summarizes the distributions considered in this subsection and the distance $D(f_{\mathcal{H}_1}(\varepsilon))$ between the null and the alternative hypotheses.

Table 3.5. 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	0.180	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	0.880	0.075	0.345	0.505

Power of the test $L^*_{\widehat{\varepsilon}}$ 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]

Unfortunately, there is no single measure that is able to quantify the empirical impact of the use of the algorithm-based estimates in conducting the hypothesis testing. In Tables 3.5 and 3.6, we quantify such an impact using the quantities in the square brackets, which are computed as the following. Hereafter, let r represent the rth replication of the 200, i.e. $r = 1, \ldots, 200$. Let $Q = \{r|\text{the replication that is rejected by } L^*_\varepsilon\}$. Similarly, let $\widehat{Q} = \{r|\text{the replication that is rejected by } L^*_{\widehat{\varepsilon}}\}$. Therefore, $\widehat{Q} \cap Q$ is the set of those replications that are rejected by both L^*_ε and $L^*_{\widehat{\varepsilon}}$. The values in the square brackets are then computed as the quotient $\frac{R_1}{R_2}$, where R_1 and R_2 are the numbers of elements in $\widehat{Q} \cap Q$ and $\widehat{Q}$, respectively. Hence, intuitively the quantities in the square brackets represent the proportion of the correct rejections made by $L^*_{\hat{\varepsilon}}$. When the quantities inside and outside of the square brackets are considered collectively, better outcomes are obtained when the differences between the power values in the first and the second panels of the tables are small and the quantities in the square brackets are close to one.

Table 3.6. 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]
$T = 501$	0.305[0.990]	0.505[0.970]	0.615[0.950]	0.090[0.777]	0.185[0.810]	0.275[0.909]

We now discuss a few important findings from Tables 3.5 and 3.6. Let us focus first on the top panel. As expected, we are able to achieve high 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 larger. Overall, the power values of L^*_{ε} look reasonable even with $\mathcal{H}_0$ and $\mathcal{H}_1$, which are quite close, and with such small numbers of observations. Now, let us concentrate on the cases where $\alpha = 5\%$, i.e. columns 2 to 4 of the tables. In all cases, the power values of the tests have a strong tendency to increase as T increases. The power values are reported to be quite close to 1 even with such a small number of observations as 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 experimental evidence in support of the consistency of the test stated in Corollary 3.2.

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

The test in this subsection is constructed so that although ε has a gamma distribution with hazard rate that is either monotonically increasing or decreasing, it may deviate locally at some finite T by a ratio of some unknown density function that converges to zero as $T \to \infty$. Specifically, the null and alternative hypotheses in this example are

$$\mathcal{H}_0 : f(\varepsilon) = f_G(\varepsilon, \theta_0) \ vs. \ \mathcal{H}_1 : f_T(\varepsilon) = (1 - C_T) f_G(\varepsilon, \theta_1) + C_T \ \varphi(\varepsilon), \tag{3.23}$$

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

$$\varphi(\varepsilon) = \left(\frac{7}{10}\right) \frac{\varepsilon^{\frac{4}{10}}}{\Gamma(2)} \exp\left\{-\varepsilon^{\frac{7}{10}}\right\}, \tag{3.24}$$

which is an inverted U-shaped hazard rate function of the generalized gamma density with $\alpha = 2$, $\delta = 0.7$ and $\beta = 1$.

Table 3.7. 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	$D(f_{\mathcal{H}_1}(\varepsilon))$	T	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$
$C_{176} = 0.096$	0.00037	176	0.335	0.125	0.285[0.835]	0.115[0.800]
$C_{301} = 0.076$	0.00026	301	0.355	0.095	0.335[0.915]	0.085[0.842]
$C_{501} = 0.051$	0.00018	501	0.360	0.125	0.340[0.864]	0.135[0.920]

The first column in Table 3.7 presents the values of probabilities, i.e. C_{176}, C_{301} and C_{501}, while the second column displays the distances between the resulting mixture density functions and the null hypothesis. Using the distance in the second column as a criterion, it is not surprising to see that the power values of the tests with the alternative hypotheses associated with C_{176}, C_{301} and C_{501} are somewhat comparable with those of SD_{C3} at $T = 176$, SD_{C2} at $T = 301$ and SD_{C1} at $T = 501$, respectively. Although the power values of 0.335, 0.355 and 0.367 in Table 3.7 seem to increase as the sample size grows, one cannot reject that it is reasonably constant. Therefore, there is a sufficient evidence to support the asymptotic results in Corollary 3.3. Furthermore, the values of $\frac{R_1}{R_2}$ in the square brackets, which are all quite close to one, are consistent with the power values of the $\widehat{\varepsilon}$-version of the tests that are approximately 90% of their ε-version counterparts.

4. Mean Residuals with Mass Close to Zero–Boundary

In many financial applications of the MEM for non-negative time series processes, often it is found that mean residuals of the series have substantial mass close to the zero-boundary. In an attempt to minimize an impact of the well–known boundary bias problem, Fernandes and Grammig (2005) follow a suggestion in the literature[15] and establish an asymmetric kernel version of the test. Their so–called D-test rests on the following distance:

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

[15] See Chen (2000), for example.

where $\int_\varepsilon$ denotes the integral over the support of the density function of ε and $I(\cdot)$ is an indicator function. The D-test gauges the discrepancy between the parametric and nonparametric estimates of the density function based on

$$\Phi_{\widetilde{f}} = \int_\varepsilon I(\varepsilon \in \mathcal{S})[f(\varepsilon, \theta_{\widetilde{f}}) - \widetilde{f}(\varepsilon)]^2 \, dF_T(\varepsilon), \tag{4.1}$$

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

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

where b_T is the bandwidth, and show that

$$L^{FG}(b_T) = \frac{T b_T^{\frac{1}{4}} \Phi_{\widetilde{f}} - b_T^{\frac{1}{4}} \widetilde{\delta}}{\widetilde{\sigma}} \underset{D}{\rightarrow} N(0,1), \tag{4.2}$$

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 undersmoothing-adjusted version of the theoretically optimal bandwidth, $\widehat{b}_T = \frac{1}{\log T}\left\{\frac{\widehat{\lambda}}{4}\right\}^{-\frac{1}{5}} (2 - \widehat{\lambda})^{-\frac{4}{5}} T^{-\frac{4}{9}}$, where $\widehat{\lambda}$ is a consistent estimator of the exponential parameter λ.

A popular alternative in the literature is to employ a boundary correction estimator. A well-known example is the suggestion of Jones (1993), which is to use a generalized jackknifing idea to obtain the $O(h^2)$ bias in both the interior and the boundary. Let

$$a_l(p) = \int_{-1}^{\min\{p,1\}} u^l k(u) du,$$

where $p = \varepsilon/h$. The idea is to take a linear combination of k with another function ℓ, which is closely related to k. A popular choice for ℓ is to set $\ell(u) = u \cdot k(u)$ and hence there is a linear relationship between $\ell(\cdot)$ and $k(\cdot)$. Such a choice results in a simple boundary corrected kernel of the form:

$$(l_u + m_u u)k(u), \quad \text{where } l_u = \tfrac{a_2(p)}{a_2(p)a_0(p) - a_1^2(p)} \text{ and } m_x = \tfrac{-a_1(p)}{a_2(p)a_0(p) - a_1^2(p)}.$$

Although there are several boundary corrected estimators using generalized jackknifing techniques, there are little differences in the performance.[16] Hence, Jones (1993) suggests

[16] Jones (1993).

that the above linear boundary corrected estimator should be an excellent choice. The resulting boundary corrected kernel density estimator can therefore be defined as follows:

$$\widetilde{f}(\varepsilon;\boldsymbol{\varepsilon},h)=\frac{1}{Th}\sum_{t=1}^{T}k_\ell\left\{\frac{\varepsilon-\varepsilon_t}{h}\right\},$$

where k_ℓ is the kernel function

$$k_\ell(u)=\frac{(a_2(p)-a_1(p)u)k(u)}{a_2(p)a_0(p)-a_1^2(p)}. \tag{4.3}$$

In practice, it is not decisive on the superiority of one of these methods compared to another. Within the context of the research in this paper, however, the statistical equivalence of (4.3) to the kernel weighted local linear fitting at the boundary suggests that the simple boundary correction estimator should provide the most straightforward extension to the testing procedure introduced in Section 2 in order to minimize the impact of the boundary bias problem. We begin by replacing $\widehat{f}(\widehat{\varepsilon}_t)$ in (2.11) with $\widetilde{f}(\widehat{\varepsilon};\widehat{\boldsymbol{\varepsilon}},h)$ to obtain the bias corrected version

$$\widetilde{N}_{T,\widehat{\varepsilon}}(h)=\frac{Th}{T}\sum_{t=1}^{T}\left(\widetilde{f}(\widehat{\varepsilon};\widehat{\boldsymbol{\varepsilon}},h)-\widetilde{f}(\widehat{\varepsilon}_t,\widetilde{\theta})\right)^2. \tag{4.4}$$

In theory, we may follow the adaptive testing procedure discussed in Section 2.2 to obtain

$$L_{\widehat{\varepsilon}}^{**}\equiv\max_{h\in H_T}\widetilde{L}_{T,\widehat{\varepsilon}}(h)=\max_{h\in H_T}\frac{\widetilde{N}_{T,\widehat{\varepsilon}}(h)-\widehat{\mu}_{T,\widehat{\varepsilon}}(h)}{\sqrt{h}\widehat{\sigma}_{T,\widehat{\varepsilon}}(h)}, \tag{4.5}$$

where $\widehat{\mu}_{T,\widehat{\varepsilon}}$, and $\widehat{\sigma}_{T,\widehat{\varepsilon}}$ are as defined previously.

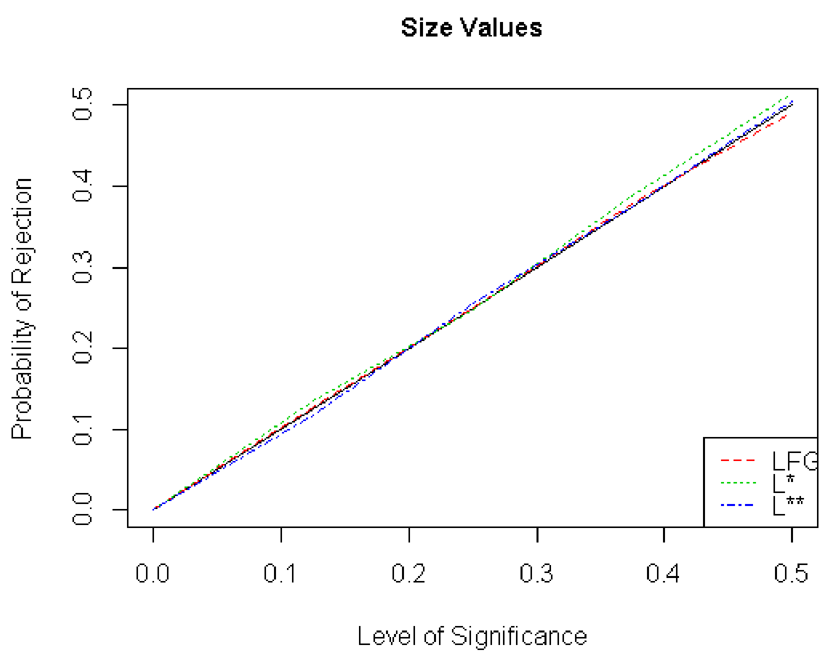
In the remainder of this section, we conduct a Monte Carlo exercise to assess the finite–sample performance of our tests, namely L^* and L^{**}, and to compare to that of the L^{FG}. The model example used in this case is the ER-ACD(1,1) model of the form

$$\psi_t=\omega+\alpha x_{t-1}+\beta\psi_{t-1}, \tag{4.6}$$

where $\alpha=0.1$, $\beta=0.7$ and $\omega=1-(\alpha+\beta)$, which was used by Fernandes and Grammig (2005).

We generate 1500 realizations of the ER-ACD(1,1) model by drawing the errors $\varepsilon_t=x_t/\psi_t$ from three distributions, namely the exponential (with $\lambda_0=1$), the gamma (with $\alpha_0=1.1$ and $\beta_0=1$) and the Weibull (with $\delta_0=1.1$ and $\beta_0=1$). The number of simulations used is 1000 replications. For each replication and data generating process, we first obtain the QML estimate of the conditional duration using the exponential distribution. The empirical estimate of the standardized duration is then computed and used in order to evaluate the finite sample performance of the tests. With regard to the bandwidth parameter, the L^{FG} test is conducted based on $\widehat{b}_T$ as suggested by Fernandes and Grammig (2005), while the L^* test is implemented adaptively as explained in the simulation in Section 3.2. With regard to the L^{**} test, we initially attempted to perform the test adaptively as done with L^*, but at the current stage the computation of this procedure in R consumed a large amount of time. Hence, the implementation of the L^{**} test in this section is based on a cross-validation MLE fitting of the boundary corrected kernel density estimation. One benefit of this implementation is the fact that the L^{**} test, in this case, can be more directly compared with the L^{FG} test of Fernandes and Grammig (2005) as a way of dealing with the mean residuals with mass close to zero-boundary. Moreover, the simulation scheme used in Section 3.2 is also performed here at $Q = 2000$ in order to compute the corresponding simulated critical values, l^*_α, l^{**}_α and l^{FG}_α at the α significance level of L^*, L^{**} and L^{FG}, respectively. The frequency of rejection of the null hypothesis is then computed in order to evaluate the size and power of the three tests. The sizes of the tests are computed based on the data simulated under $\mathcal{H}_0$ of the above exponential, gamma and Weibull distributions. Conversely, the power values of the tests are calculated based on the data generated under $\mathcal{H}_1$. In particular, the estimated model specifies an exponential distribution whereas the true density belongs to either the gamma or the Weibull distribution. Finally, in order to compute the nonparametric estimators involved, we choose the normal kernel function.

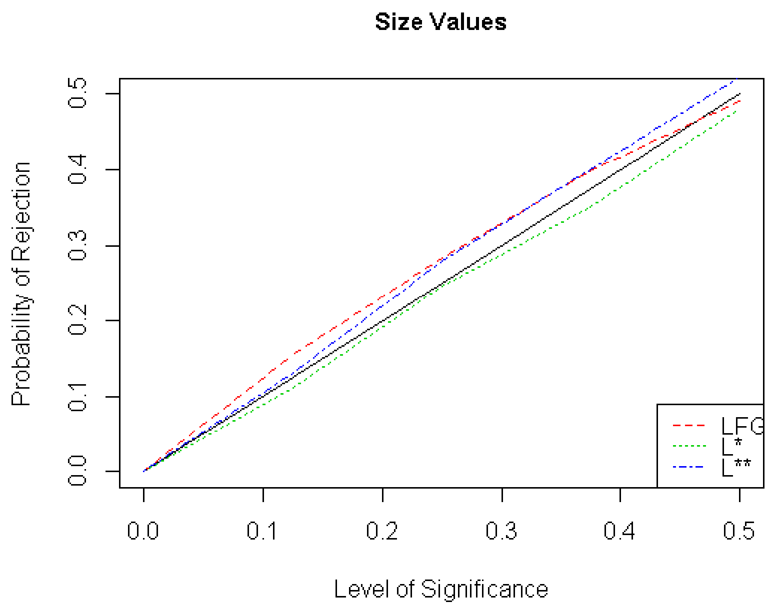
Figure 4.1. *Data generating process: Exponential ER-ACD(1,1).*



Figures 4.1 to 4.3 display the main results using Fernandes and Grammig (2005) type of graphical representation. While Figure 4.1 presents the sizes of the tests at different levels of significance (displayed in the figure as the dark-solid line) based on an exponential model, the top panels of Figures 4.2 and 4.3 present those of the gamma and the Weibull models, respectively. In this aspect, it seems that all the three tests perform reasonably well, while the L^* test slightly outperforms the remaining two tests with regard to the gamma and Weibull models. Furthermore, the bottom panels of Figures 4.2 and 4.3 present power

values based the gamma and Weibull models at different levels of significance, respectively. The null hypothesis in these cases is the exponential distribution. In this aspect, the L^* test clearly outperforms the remaining two tests, especially with regard to the Weibull model, while the L^{**} test performs slightly better than L^{FG}.

Figure 4.2. *Data generating process: Gamma ER-ACD(1,1).*



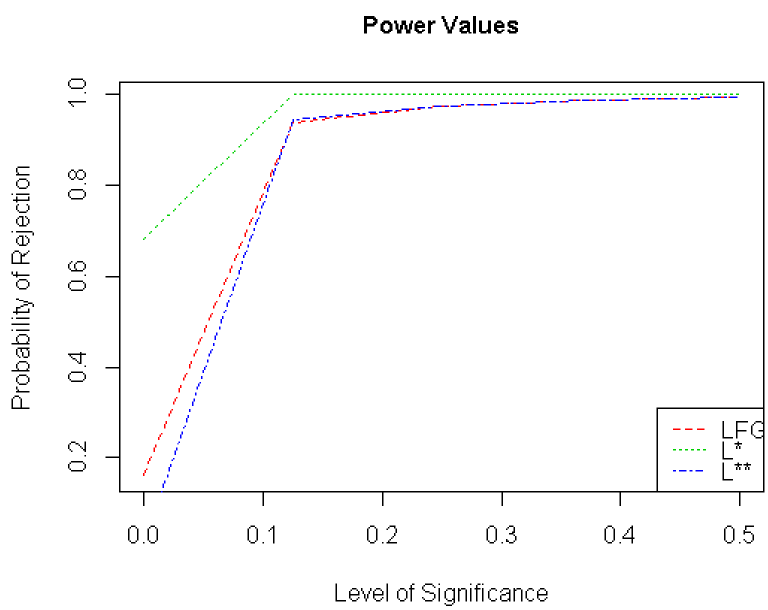
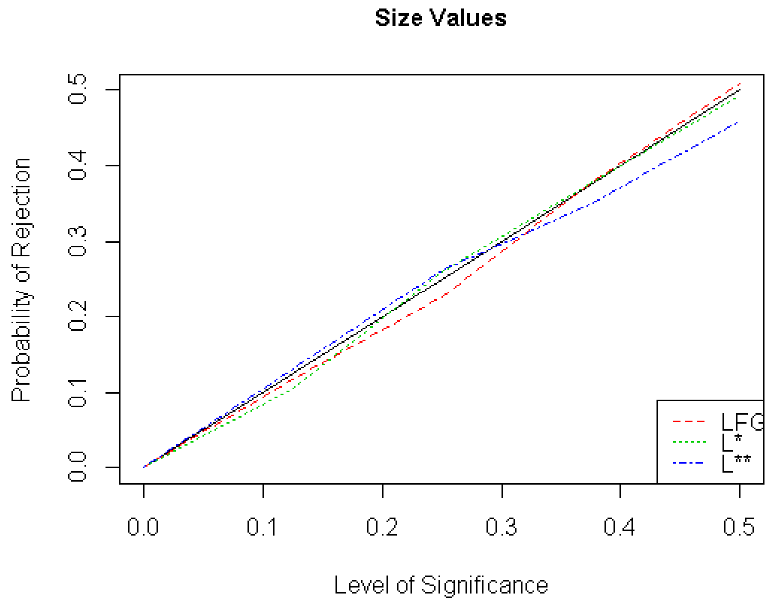
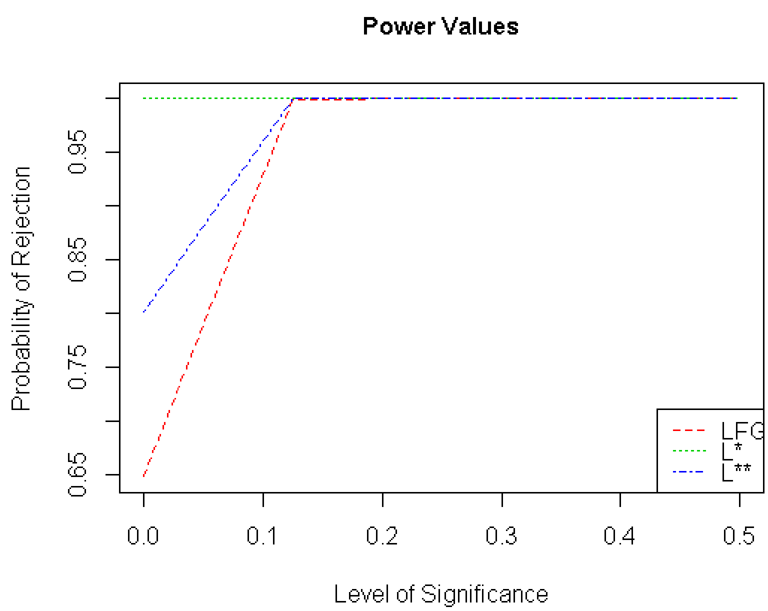


Figure 4.3. *Data generating process: Weibull ER-ACD(1,1).*





5. Econometric Analysis of the Price Duration Process

In this section, we conduct an empirical analysis of the price duration processes at the NYSE and the ASX using the models and the adaptive test procedure presented in the previous sections. The analysis in this section focuses on two separate issues with each being studied based on a different set of data as the following. (1) Firstly, it is the IBM data used in Engle and Russell (1998). A total of 60,328 transactions are recorded for IBM over the three months of trading on the consolidated market from November 1990 through to January 1991. As per the seminal paper, two days from the three months were deleted. A halt occurred on 23 November and an opening delay of more than one hour occurred on 27 December. (2) Secondly, it is tick-by-tick data of AMP stock, which is listed on the ASX for the period of April to June 2000. The first half hour of the trading days, i.e. trades before 10:00 a.m., 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. Finally, all trades after 04:00 p.m. are also omitted.

In order to obtain the price duration process, we first perform data thinning so that in essence only the points at which the price has changed significantly since the occurrence of the last price change are kept. To minimize the effects of errant quotes, two consecutive points were required to have changed by more than a threshold value, c, since the last price change. A more formal explanation of the thinning method can be given 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 (e.g. $j = 1$ for the starting stage of the thinning process) and the constant c represents a threshold value. The choice of c depends upon a number of specific characteristics of the financial market in question. For example, Engle and Russell (1997) use $c = 0.0005$ when conducting an analysis on the price duration in the foreign exchange market, while Engle and Russell (1998) set $c = 0.25$ for a similar study on the NYSE. In this paper, we follow the latter and set $c = 0.25$.

Table 5.1. Descriptive statistics

Descriptive Statistics	ν_{IBM}	$\tilde{x}_{IBM}$	$\widehat{\varepsilon}_{IBM}$	ν_{AMP}	$\tilde{x}_{AMP}$	$\widehat{\varepsilon}_{AMP}$
Mean	860.00	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 Table 5.1 present the descriptive statistics of IBM and AMP's observed price duration, respectively. While the average price durations for the IBM and AMP samples are 860 and 900 seconds, the maxima are 10,609 and 15,082 seconds, respectively. The minimum price duration for both IBM and AMP is 1 second. Finally, the Ljung-Box test values of 42.50 and 91.60 indicate strong clustering behavior and autocorrelation in both the IBM and AMP price duration series.

It is widely documented in the literature[17] that price durations feature a strong time-of-day effect, which is related to predetermined market characteristics, e.g. trade open-

[17]See Giot (2000), and Meitz and Teräsvirta (2006), for example.

ing/closing times and lunch hours. In the current paper, we assume that the stationary price duration series can be computed as

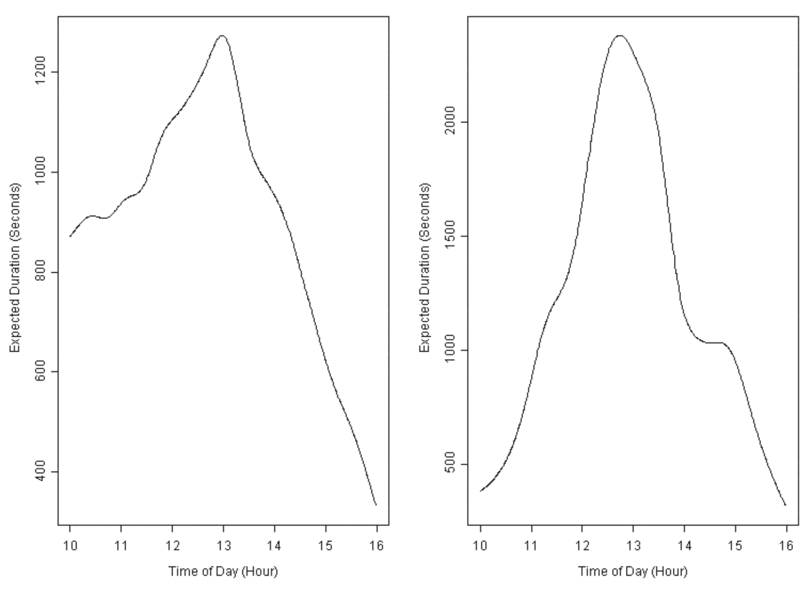
$$x_t = \frac{\nu_t}{\phi(i_{t-1})} = \psi_t \varepsilon_t, \tag{5.1}$$

where ν_t denotes the observed price duration as constructed in the previous paragraph and $\phi(i_{t-1})$ denotes an intraday diurnal factor. Following Saart et al. (2013), we estimate the diurnal factor $\phi(i_{t-1})$ of the calendar time i_{t-1} at which the tth 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 $W_{v,h}(y) = K_h(y - i_{t-1})/\sum_{t=2}^{T} K_h(y - i_{t-1})$. In our calculation, an asymptotically optimal bandwidth parameter is selected using the leave-one-out cross validation selection criterion. Figure 5.1 presents the kernel estimates of the diurnal factors associated with the IBM (left-panel) and AMP (right-panel) price duration processes. As expected, the price durations are shortest in the morning and just prior to the close, with a noticeable lull during lunch hours. These results are consistent with those found in existing literature. [18]

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

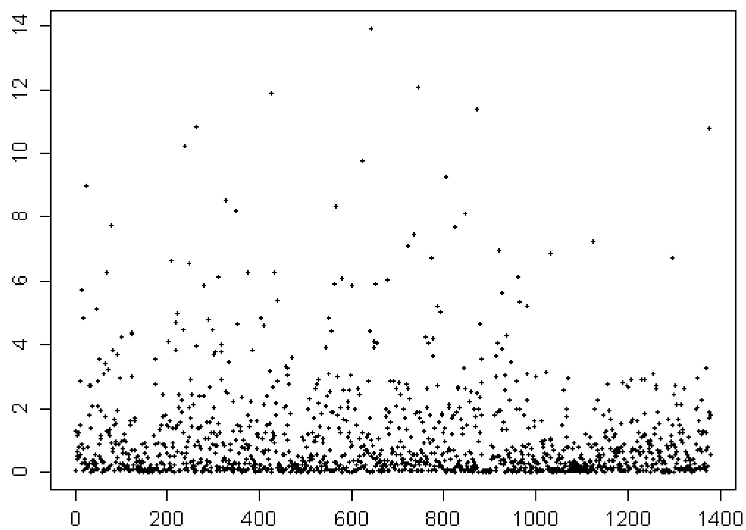


The next step is to model the ratio of the actual to the fitted value $\widetilde{x}_t = \nu_t/\widehat{\phi}_h(i_{t-1})$ as an ACD model of the diurnally adjusted series of price durations. However, before doing so,

[18]See Engle and Russell (1998), for example.

let us check to see if the diurnal adjustment alone is able to take care of the serial correlation and the clustering behavior of the duration processes. The Ljung–Box statistics reported in the third and the sixth columns of Table 5.1 suggest that the diurnally adjusted price duration series of both IBM and AMP still exhibit these peculiar time series features. Hence, the use of the ACD-procedure to model the stochastic component of price duration processes is essential.

Figure 5.2. *Standardized duration series for IBM.*



5.1. Analysis of the Price Duration Process for IBM Data.

When conducting an analysis using this dataset, Engle and Russell (1998) employed the ER-ACD(1,1) model based on the Weibull distribution. In this section, we take the view that the SEMI-ACD model that ensures flexibility not only in the error distribution (since the distributional assumption of the error term is not required), but also in the conditional duration and, therefore, should be the most suitable in minimizing the impact of a potential joint hypothesis problem in the MEM. Hence, our modeling procedure in this section is first to apply the SEMI-ACD(1,1) model to obtain the empirical estimate of the standardized duration, then use the above-discussed adaptive test procedure to test misspecification of the Weibull distribution.[19] To estimate the SEMI–ACD model, we employ the quartic kernel function $K(u) = \left(\frac{15}{16}\right)(1-u^2)^2$. An estimation-based optimal bandwidth for each of the iteration steps is selected using an adaptive cross-validation algorithm. Finally, we select $m^* = 7$ by using a similar procedure to what has been discussed in Saart et al. (2013).

[19]See Appendix B.1 for a brief review and Saart et al. (2013) for more detailed discussion on the theory and practice of the SEMI-ACD Model.

The empirical estimate of the standardized duration can be computed using the formula:

$$\widehat{\varepsilon}_{t,m^*} = \frac{\nu_t}{\widehat{\phi}_h(i_{t-1})\widehat{\psi}_{t,m^*}}. \tag{5.3}$$

The descriptive statistics of the series are presented in Table 5.1. Figures 5.2 and 5.3 present the standardized duration series and the kernel density estimate of the probability density function, respectively.

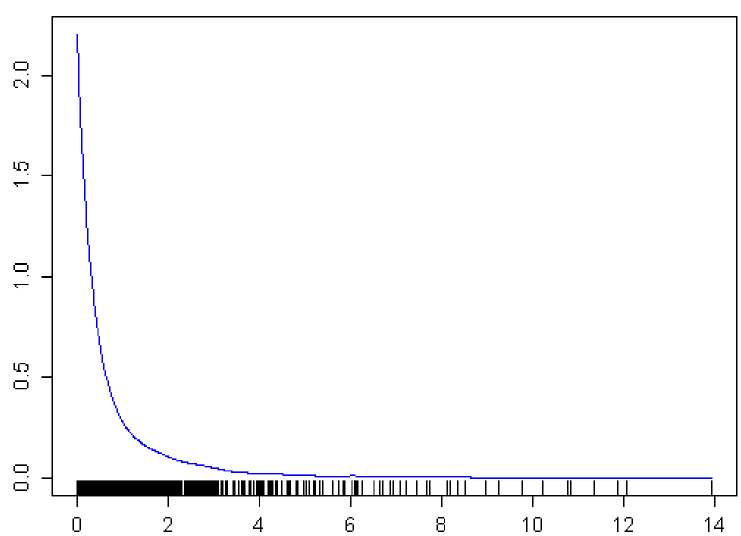
Maximum likelihood estimates of the three most often associated with models of financial duration are the exponential with the rate of 0.9797, gamma with the shape and scale parameters of 0.5828 and 1.4285, respectively, and the Weibull with the shape and scale parameters of 0.6914 and 0.7852, respectively. In what follows, we attempt to shed further light on the distribution properties of the standardized durations by employing our newly developed test procedure. We propose the following steps for computing the p-values of $L^*_{\widehat{\varepsilon}}$:

Step 5.1.1: Compute $\{\widehat{\psi}_{t,m^*}\}$ and generate $\{\widehat{\varepsilon}^*_t\}$, which is a sequence of i.i.d. bootstrap re–samples under the null hypothesis.

Step 5.1.2: Compute $\widehat{x}^*_t = \widehat{\psi}_{t,m^*}\widehat{\varepsilon}^*_t$, then use it in the computation of $L^*_{\widehat{\varepsilon}}$.

Step 5.1.3: Repeat the proceeding steps Q times in order to produce Q versions of $L^*_{\widehat{\varepsilon}}$, i.e. $L^*_{\widehat{\varepsilon},q}$ for $q = 1, 2, \cdots, Q$. Find the bootstrap distribution of $L^*_{\widehat{\varepsilon},q}$ and then compute the proportion in which $L^*_{\widehat{\varepsilon}} < L^*_{\widehat{\varepsilon},q}$. This proportion is then a simulated p-value of $L^*_{\widehat{\varepsilon}}$.

Figure 5.3. *Kernel density estimate of the probability density function.*



The use of the maximized version of the test statistic suggests that the selection of H_T can have significant impacts on the final conclusions of the results. For model estimation and hypothesis testing, the specific formulation of H_T employed here is a geometric grid of the form $H_T = \{h = h_{\max} a^k : h \geq h_{\min},\ k = 0, 1, 2, \ldots\}$, where $0 < h_{\min} < h_{\max}$ and $0 < a < 1$, with $J_T \leq \log_{1/a}(h_{\max}/h_{\min})$ representing the number of elements in H_T. In order to perform the hypothesis tests, we choose $a = 0.25$ and $b = 0.4$.

In order to conduct a misspecification test of the Weibull distribution in Engle and Russell (1998), we test the following hypotheses:

(i) $\mathcal{H}_0 : \exists \theta_0 \in \Theta$ such that $f(\varepsilon) = f_W(\varepsilon, \theta_0)$ against the alternative hypothesis that there is no such $\theta_0 \in \Theta$.

If this is rejected, then we perform the following test:

(ii) $\mathcal{H}_0 : \exists \theta_0 \in \Theta$ 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. Hypothesis testing on the parametric density function of the standardized duration.

Hypotheses	Distributions	p-values
(i)	Weibull	0.0189
(ii)	Gamma	0.0205
(iii)	Generalized Gamma	0.1295

Note: p-values are based on the empirical distribution of the test statistic stemming from 2000 artificial bootstrap samples.

Figure 5.4. *Kernel density estimate of the density function.*

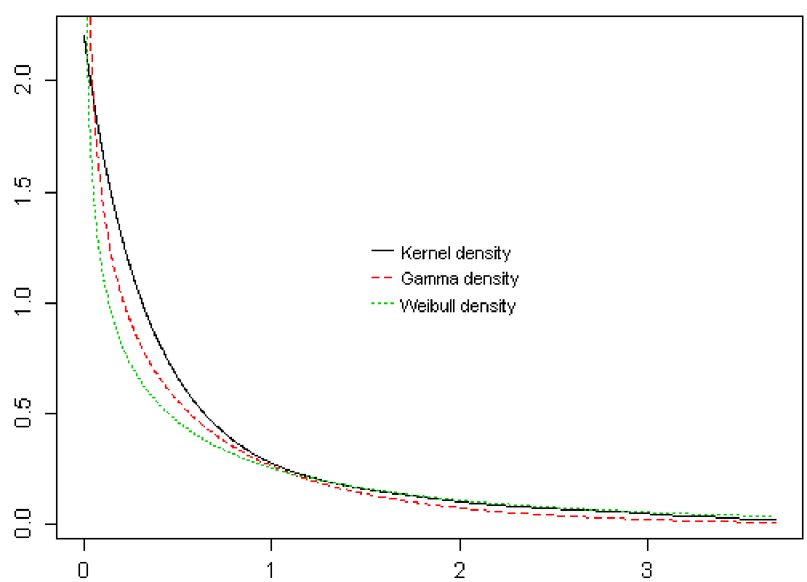
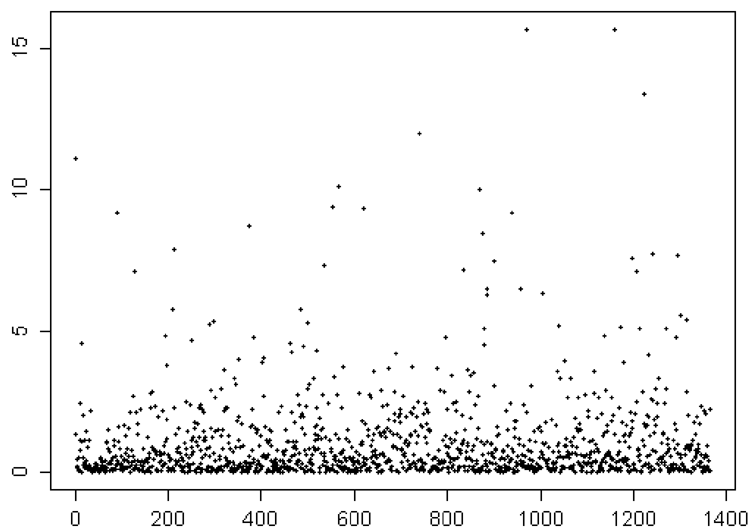


Table 5.2 reports the results of our hypothesis tests. In the table, the resulting p-value suggests that the null hypothesis of the Weibull distribution is rejected at the 5% level of significance. This is a strong evidence for a misspecification of the WACD model. Such a misspecification is anticipated by Engle and Russell (1998), but an evidence is not formally provided. Furthermore, the null hypothesis of the gamma distribution is also rejected at the 5% level of significance. Hence, further analysis is clearly required for IBM.

Figure 5.4 presents three estimates of the probability density function of the standardized duration, namely the nonparametric kernel, the gamma (with $\alpha = 0.75$ and $\theta = 1$) and the Weibull (with $\theta = 1$ and $\delta = 0.75$). By taking the view that the kernel density estimator is consistent, it is clear that the gamma density function seems to fit better for values of the standardized durations below 1. On the other hand, the Weibull density function seems to fit better for larger values. This provides graphical evidence that points toward a more general distribution that nests both the gamma and the Weibull distributions as special cases. This leads us to the third testing strategy, which is to test the null hypothesis of a generalized gamma distribution. The results in Table 5.2 show that the null hypothesis of the three-parameter generalized gamma distribution is not rejected at the usual 5% level of significance.

Figure 5.5. *Standardized duration series for AMP.*



5.2. Analysis of the Price Duration Process for AMP Data.

In Section 1, we briefly introduce the idea of treating the Engle and Russell's (1998) ACD framework as a duration-based time-scaling instrument since such a concept is well-known for the intensity-based approaches of point processes. The empirical study in this section investigates the performance of the ACD framework in performing this task in practice. The study is conducted based on the price duration process of AMP that we computed at the beginning of this section and a set of modeling steps, which can be described as follows:

Step 5.2.1: Obtain the QML estimate of the ER-ACD(1,1) conditional duration, i.e.

$$\psi_t = \alpha x_{t-1} + \beta \psi_{t-1},$$

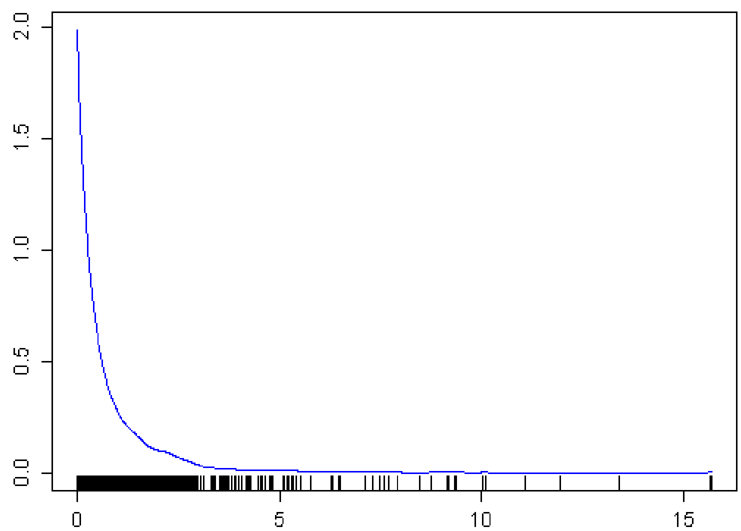
where x_t is as defined in (5.1), using the exponential distribution.

Step 5.2.2: Compute the standardized duration as in (5.3) based on the estimate obtained in Step 5.2.1.

Step 5.2.3: Perform the relevant hypothesis testing based on $L_{\hat{\varepsilon}}^{**}$ and draw the conclusion.

Performing Step 5.2.1 is straightforward and is frequently discussed in the literature, so its discussion is omitted here. Figures 5.5 and 5.6 present the resulting standardized duration series and the kernel density estimate of the probability density function, respectively.

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



In this section, we take the view that for the Engle and Russell's (1998) ACD framework to perform well, as a duration-based time-scaling instrument, the resulting standardized duration must be exponentially distributed (equivalently, the arrival process that can be derived based on such standardized duration must follow a Poisson process). Therefore, we will perform the test of the following hypotheses:

(iv) $\mathcal{H}_0 : \exists\lambda_0 \in \Theta$ such that $f(\varepsilon) = f_E(\varepsilon, \lambda_0)$ against the alternative hypothesis that there is no such $\lambda_0 \in \Theta$.

The p-value for value for the test is computed using the simulation scheme introduced in Steps 5.1.1 to 5.1.3.

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

Hypotheses	Distributions	p-values
(iv)	Exponential	0.0000

Note: p-values are based on the empirical distribution of the test statistic stemming from 2000 artificial bootstrap samples.

Table 5.3 reports the result of our hypothesis test. The p-value suggests that the null hypothesis of the exponential distribution should be rejected at the 5% level of significance. Hence, the ACD framework is not able to perform well as a duration-based time-scaling instrument. We believe that a potential source of the problem may be in the modelling of the self-excited dynamics of the duration process, which is done through the specification of the conditional duration in the ACD framework.

6. Conclusions

In this paper, we have considered misspecification testing for the MEM of non-negative time series processes. Particularly, we have presented an alternative testing procedure to test the probability distribution of the error term in the MEM. The misspecification in the distribution of the error term has a number of important statistical and practical implications, especially within the context of the MEM of the duration process in finance. Our testing procedure differed from those that appeared in previous studies since it was established based on an intricate network of recent developments in semi/nonparametric literature. In this paper, we established and discussed the asymptotic distribution of the test statistic, and showed theoretically and experimentally the consistency of the adaptive test procedure employed. Finally, we have applied our testing procedure to the MEM of price duration processes using data sets from the NYSE and the ASX. Overall, we believe that this paper sheds some new light into statistical inferences of the MEM and the semi- and non-parametric models with generated covariates in general.

7. Appendix A

This Appendix consists of three subsections. Section A.1 presents a set of assumptions, including the conditions on both $h_{\min}$ and $h_{\max}$. While Section A.2 discusses a number of technical Lemmas, Section A.3 gives the proofs of the main results established in Section 2.

A.1. Technical Assumptions:

Assumption A.1. *Let Assumptions 2.1 and 2.2 hold.*

Assumption A.2. *The bandwidth h satisfies* $\lim_{T\to\infty} h = 0$, $\lim_{T\to\infty} Th^2 = \infty$ *and* $\lim\sup_{T\to\infty} Th^5 < \infty$.

Assumption A.3. *(i) Let Assumption 2.3 holds.*

(ii) Let H_T be specified as in (2.18) with $c_{\min}T^{-\gamma} = h_{\min} < h_{\max}$, where $h_{\max} = c_{\max}\left(\text{loglog}(T)\right)^{-1}$, and γ, $c_{\min}$, $c_{\max}$ are some constants satisfying $0 < \gamma < 1$ and $0 < c_{\min}, c_{\max} < \infty$.

A.2. Technical Lemmas:

Lemma A.1. *Let Assumptions A.1 and A.2 hold. In addition, let*

$$\Delta_{k_h}(t) \equiv \Delta_{k_h}(\varepsilon_t, \varepsilon_t; h) = \frac{1}{T}\sum_{s=1}^{T} k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - \frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s).$$

Then, we have, as $T \to \infty$,

$$\sup_{h\in H_T}\max_{1\le t\le T} |\Delta_{k_h}(t)| = o_P\left\{T^{-\delta}\right\}. \tag{A.1}$$

Proof. Let us first consider the term $\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - k_h(\varepsilon_t - \varepsilon_s)\}$. A Taylor expansion of k_h of order two at $\{\varepsilon_t - \varepsilon_s\}$ suggests that:

$$k_h[\widehat{\varepsilon}_t - \widehat{\varepsilon}_s] \approx k_h[\varepsilon_t - \varepsilon_s] + k_h^{(1)}[\varepsilon_t - \varepsilon_s]\Delta_{ts} + \frac{1}{2}k_h^{(2)}[\varepsilon_t - \varepsilon_s]\Delta_{ts}^2, \tag{A.2}$$

where $\Delta_{ts} = \{\widehat{\varepsilon}_t - \widehat{\varepsilon}_s\} - \{\varepsilon_t - \varepsilon_s\} = \{\widehat{\varepsilon}_t - \varepsilon_t\} - \{\widehat{\varepsilon}_s - \varepsilon_s\}$. This further suggests that

$$k_h[\widehat{\varepsilon}_t - \widehat{\varepsilon}_s] - k_h[\varepsilon_t - \varepsilon_s] \approx \sum_{q=1}^{2} k_h^{(q)}[\varepsilon_t - \varepsilon_s]\{\widehat{\varepsilon}_t - \varepsilon_t\}^q + \sum_{q=1}^{2} k_h^{(q)}[\varepsilon_t - \varepsilon_s]\{\widehat{\varepsilon}_s - \varepsilon_s\}^q$$

$$+ k_h^{(2)}[\varepsilon_t - \varepsilon_s]\left\{\{\widehat{\varepsilon}_t - \varepsilon_t\}\{\widehat{\varepsilon}_s - \varepsilon_s\}\right\} + R\{\Delta_{ts}; \varepsilon_t - \varepsilon_s\}, \tag{A.3}$$

where there exists c^* between $\{\varepsilon_t - \varepsilon_s\}$ and $\{\varepsilon_t - \varepsilon_s\} + \Delta_{ts}$ such that $R\{\Delta_{ts}; \varepsilon_t - \varepsilon_s\} = \frac{1}{3!}k_h^{(3)}[c^*]\Delta_{ts}^3$ denotes the remainder by which $k_h^{(r)}[\cdot] = \left(\frac{1}{h^{i+1}}\right) k^{(r)}\left(\frac{\cdot}{h}\right)$ for any integer $r \geq 1$.

Below the main idea is to use the α–mixing condition to replace the joint density function, $p_\tau(x, y)$, of $(\varepsilon_t, \varepsilon_s)$ by the product of the marginal density functions $p(x)p(y)$, when $|t - s| = \tau$ is large enough. Since the mixing coefficient of $\{\varepsilon_t\}$ decays exponentially, we need only to replace $p_\tau(x, y)$ by $p(x)p(y)$ in the following derivations (see, for example, Appendix A of Gao (2007)).

Observe that we have:

$$\frac{1}{Th^{1+r}} \sum_{s=1}^{T} E\left[\left|k^{(r)}\left[\frac{\varepsilon_t - \varepsilon_s}{h}\right]\right|\right] = \frac{1}{h^{1+r}} \int\int \left|k^{(r)}\left[\frac{v-u}{h}\right]\right| f(u, v)\ du\ dv$$
$$= (1 + o(1))\frac{1}{h^{1+r}} \int\int \left|k^{(r)}\left[\frac{v-u}{h}\right]\right| f(u)f(v)\ du\ dv. \tag{A.4}$$

We know from Assumption 2.2(iii) that $k^{(q)}(u)$ is bounded. In the proof that follows, we need to show only that (A.4) is asymptotically bounded. Therefore, in order to avoid any unnecessary complication and without loss of generality, let us proceed by taking integration by parts and changes of variables of

$$\frac{1}{h^{1+r}} \int\int k^{(r)}\left[\frac{v-u}{h}\right]\ f(u)f(v)\ du\ dv \tag{A.5}$$

to obtain

$$\frac{1}{h^{1+r}} \int\int k^{(r)}\left[\frac{v-u}{h}\right] f(u)f(v)du\,dv = -\frac{1}{h} \int\int k\left[\frac{v-u}{h}\right] f(u)f^{(r)}(v)du\,dv$$
$$= \int\int k[u]f(v - hu)f^{(r)}(v)(dv)du\,dv = \int k[u]du \int f(v)f^{(r)}(v)dv + O(h^p), \tag{A.6}$$

where the final equality is by a pth-order Taylor series expansion. Note also that we have

$$\frac{1}{h}E\left[\left|k\left[\frac{\varepsilon_t - \varepsilon_s}{h}\right]\right|\right] = \frac{1}{h}\int\int \left|k\left[\frac{v-u}{h}\right]\right| f(u, v)du\,dv$$
$$= (1 + o(1))\frac{1}{h}\int\int \left|k\left[\frac{v-u}{h}\right]\right| f(u)f(v)du\,dv = (1 + o(1)) \int\int |k[u]|\ \ f(v - uh)f(v)(dv)\,du\,dv$$
$$= (1 + o(1))\left\{\int |k[u]|\,du \cdot \int (f(v))^2\ (dv)\,dv + O(h^p)\right\}. \tag{A.7}$$

We now use the fact that $\Delta_{k_h}(t) = \frac{1}{T}\sum_{s=1}^{T}\left\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - k_h(\varepsilon_t - \varepsilon_s)\right\}$, the above Taylor approximation and (A.6). Therefore, Assumption 2.1 and the standard ergodic theorem collectively imply that we have for $q = 1, 2$

$$\sup_{h\in H_T}\max_{1\le t\le T}|\Delta_{k_h,1q}(t)| \equiv \sup_{h\in H_T}\max_{1\le t\le T}\left|\frac{1}{T}\sum_{s=1}^{T}k_h^{(q)}(\varepsilon_t-\varepsilon_s)(\widehat{\varepsilon}_t-\varepsilon_t)^q\right|$$

$$\le \sup_{h\in H_T}\max_{1\le t\le T}|\widehat{\varepsilon}_t-\varepsilon_t|^q\left\{\frac{1}{T}\sum_{s=1}^{T}|k_h^{(q)}(\varepsilon_t-\varepsilon_s)|\right\} = o_P\{T^{-\delta}\}^q, \tag{A.8}$$

$$\sup_{h\in H_T}\max_{1\le t\le T}|\Delta_{k_h,2q}(t)| \equiv \sup_{h\in H_T}\max_{1\le s\le T}\left|\frac{1}{T}\sum_{s=1}^{T}k_h^{(q)}(\varepsilon_t-\varepsilon_s)(\widehat{\varepsilon}_s-\varepsilon_s)^q\right|$$

$$\le \sup_{h\in H_T}\max_{1\le s\le T}|\widehat{\varepsilon}_s-\varepsilon_s|^q\left\{\frac{1}{T}\sum_{s=1}^{T}|k_h^{(q)}(\varepsilon_t-\varepsilon_s)|\right\} = o_P\{T^{-\delta}\}^q, \tag{A.9}$$

$$\sup_{h\in H_T}\max_{1\le s,t\le T}|\Delta_{k_h,3}(t)| \equiv \sup_{h\in H_T}\max_{1\le s,t\le T}\left|\frac{1}{T}\sum_{s=1}^{T}k_h^{(2)}(\varepsilon_t-\varepsilon_s)(\widehat{\varepsilon}_s-\varepsilon_s)(\widehat{\varepsilon}_t-\varepsilon_t)\right|$$

$$\le \sup_{h\in H_T}\max_{1\le s,t\le T}\{|\widehat{\varepsilon}_s-\varepsilon_s|\ |\widehat{\varepsilon}_t-\varepsilon_t|\}\left\{\frac{1}{T}\sum_{s=1}^{T}|k_h^{(2)}(\varepsilon_t-\varepsilon_s)|\right\} = o_P\{T^{-\delta}\}^2. \tag{A.10}$$

Finally, equations (A.6) and (A.8) to (A.10) imply that we have:

$$\sup_{h\in H_T}\max_{1\le t\le T}|\Delta_{k_h}(t)| \le o_P\{T^{-\delta}\} \tag{A.11}$$

as claimed. ■

Lemma A.2. *Let Assumptions A.1 and A.2 hold. In addition, let*

$$s_r(\widehat{\varepsilon}_t) - s_r(\varepsilon_t) = \frac{1}{T}\sum_{s=1}^{T}k_h(\widehat{\varepsilon}_t-\widehat{\varepsilon}_s)(\widehat{\varepsilon}_t-\widehat{\varepsilon}_s)^r - \frac{1}{T}\sum_{s=1}^{T}k_h(\varepsilon_t-\varepsilon_s)(\varepsilon_t-\varepsilon_s)^r \tag{A.12}$$

for $r = 0, 1, 2$. Then we have, as $T \to \infty$,

$$\sup_{h\in H_T}\max_{1\le t\le T}|s_r(\widehat{\varepsilon}_t) - s_r(\varepsilon_t)| = o_P\{T^{-\delta}\}. \tag{A.13}$$

Proof. We start with the case of $r = 0$. Note that we have

$$\Delta_{s0}(h) = \frac{1}{T}\sum_{s=1}^{T}\left\{k_h(\widehat{\varepsilon}_t-\widehat{\varepsilon}_s) - k_h(\varepsilon_t-\varepsilon_s)\right\}. \tag{A.14}$$

Hence, it is immediately the case that

$$\sup_{h\in H_T}\max_{1\le s,t\le T}|\Delta_{s0}(h)| \le o_P\{T^{-\delta}\} \tag{A.15}$$

following (A.11). Furthermore, we have when $r = 1$

$$\begin{aligned}\Delta_{s1}(h) &= s_1(\widehat{\varepsilon}_t) - s_1(\varepsilon_t) = \frac{1}{T}\sum_{s=1}^{T} k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - \frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s)(\varepsilon_t - \varepsilon_s) \\ &= \frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s)(\widehat{\varepsilon}_t - \varepsilon_t) - \frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s)(\widehat{\varepsilon}_s - \varepsilon_s) \\ &+ \frac{1}{T}\sum_{s=1}^{T} \{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - k_h(\varepsilon_t - \varepsilon_s)\}\left[\{\widehat{\varepsilon}_t - \varepsilon_t\} + \{\varepsilon_t - \varepsilon_s\} - \{\widehat{\varepsilon}_s - \varepsilon_s\}\right].\end{aligned}$$

Taking the upper bound of $\Delta_{s1}(h)$ leads to:

$$\begin{aligned}\sup_{h\in H_T}\max_{1\le t\le T}|\Delta_{s1}(h)| &= \sup_{h\in H_T}\max_{1\le t\le T}|s_1(\widehat{\varepsilon}_t) - s_1(\varepsilon_t)| \\ &\le \sup_{h\in H_T}\max_{1\le t\le T}|s_{1,t}(h)| + \sup_{h\in H_T}\max_{1\le t\le T}|s_{1,s}(h)| + \sup_{h\in H_T}\max_{1\le t\le T}|s_{1,\Delta_{k_h}}(h)|. \qquad (A.16)\end{aligned}$$

Note that we have for $j = \{s, t\}$

$$\begin{aligned}\sup_{h\in H_T}\max_{1\le j\le T}|s_{1,j}(h)| &= \sup_{h\in H_T}\max_{1\le j\le T}\left|\frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s)(\widehat{\varepsilon}_j - \varepsilon_j)\right| \\ &\le \sup_{h\in H_T}\max_{1\le j\le T}|\widehat{\varepsilon}_j - \varepsilon_j|\left\{\frac{1}{T}\sum_{s=1}^{T}|k_h(\varepsilon_t - \varepsilon_s)|\right\} = o_P\{T^{-\delta}\}\end{aligned}$$

using (A.7), Assumption 2.1 and the standard ergodic theorem. The proof of $\sup_{h\in H_T}\max_{1\le t\le T}|s_{1,\Delta_{k_h}}(h)| = o_P\{T^{-\delta}\}$ relies on the convergence rate of $\Delta_{k_h}(h)$ in (A.11) and the fact that $E\left[|\varepsilon_t - \varepsilon_s|\right] < \infty$. Thus, we have

$$\sup_{h\in H_T}\max_{1\le t\le T}|s_{1,\Delta_{k_h}}(h)| \le \sup_{h\in H_T}\max_{1\le t\le T}|\{\widehat{\varepsilon}_t - \varepsilon_t\}\Delta_{s0}| \le o_P\{T^{-\delta}\}^2. \qquad (A.17)$$

For the case of $r = 2$, we have

$$\begin{aligned}&s_2(\widehat{\varepsilon}_t) - s_2(\varepsilon_t) = \frac{1}{T}\sum_{s=1}^{T} k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)^2 - \frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s)(\varepsilon_t - \varepsilon_s)^2 \\ &= \frac{1}{T}\sum_{s=1}^{T} k_h(\varepsilon_t - \varepsilon_s)\left\{(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)^2 + (\varepsilon_s - \varepsilon_s)^2\right\} + \frac{1}{T}\sum_{s=1}^{T}\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - k_h(\varepsilon_t - \varepsilon_s)\}(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)^2 \\ &= S_{2,1}(h) + S_{2,2}(h).\end{aligned}$$

Let $\widehat{\delta}_t \equiv \widehat{\varepsilon}_t - \varepsilon_t$, $\widehat{\delta}_s \equiv \widehat{\varepsilon}_s - \varepsilon_s$ and $\delta_{st} = \varepsilon_t - \varepsilon_s$. Observe that for $S_{2,j}(h)$ $(j = 1, 2)$,

$$\begin{aligned}\left|\{\widehat{\varepsilon}_t - \widehat{\varepsilon}_s\}^2 - \{\varepsilon_t - \varepsilon_s\}^2\right| &= \left|\{\widehat{\delta}_t + \delta_{st} - \widehat{\delta}_s\}^2 - \{\delta_{st}\}^2\right| \\ &\le \left|\widehat{\delta}_t\right|^2 + \left|\widehat{\delta}_s\right|^2 + 2\left|\widehat{\delta}_t\delta_{st}\right| + 2\left|\widehat{\delta}_s\delta_{st}\right| + 2\left|\widehat{\delta}_t\widehat{\delta}_s\right|. \qquad (A.18)\end{aligned}$$

Taking the upper bound of (A.18) leads to

$$\begin{aligned}&\sup_{h\in H_T}\max_{1\le s,t\le T}\left\{\left|\widehat{\delta}_t\right|^2 + \left|\widehat{\delta}_s\right|^2\right\} + \sup_{h\in H_T}\max_{1\le s,t\le T}\left\{\left|\widehat{\delta}_t\right| + \left|\widehat{\delta}_s\right|\right\}B_{ts} \\ &+ \sup_{h\in H_T}\max_{1\le t\le T}\left|\widehat{\delta}_t\right| \cdot \sup_{h\in H_T}\max_{1\le s\le T}\left|\widehat{\delta}_s\right| = o_P\{T^{-\delta}\},\end{aligned}$$

where $B_{st} = \max_{s,t} |\varepsilon_t - \varepsilon_s|$. Equations (A.11) and (A.7) suggest that we have $\sup_{h \in H_T} \max_{1 \leq s,t \leq T} |S_{2,j}(h)| = o_P\{T^{-\delta}\}$ for $j = 1, 2$. ■

Lemma A.3. *Let Assumptions A.1 and A.2 hold. In addition, let*

$$\widehat{R}_{T,\widehat{\varepsilon}}(h) = \sum_{t=1}^{T} (\widehat{f}(\widehat{\varepsilon}_t) - \tilde{f}(\widehat{\varepsilon}_t, \tilde{\theta}))^2 \text{ and } \widehat{R}_{T,\varepsilon}(h) = \sum_{t=1}^{T} (\widehat{f}(\varepsilon_t) - \tilde{f}(\varepsilon_t, \tilde{\theta}))^2.$$

Then we have, as $T \to \infty$,

$$\widehat{R}_{T,\widehat{\varepsilon}}(h) = \widehat{R}_{T,\varepsilon}(h) + o_P\{1\} \text{ uniformly over } h \in H_T. \tag{A.19}$$

Proof. Observe that

$$\widehat{R}_{T,\widehat{\varepsilon}}(h) = \sum_{t=1}^{T} (\widehat{f}(\widehat{\varepsilon}_t) - \tilde{f}(\widehat{\varepsilon}_t, \tilde{\theta}))^2 = \widehat{R}_{T,\widehat{\varepsilon},1}(h) + \widehat{R}_{T,\widehat{\varepsilon},2}(h), \tag{A.20}$$

where $\widehat{R}_{T,\widehat{\varepsilon},2}(h)$ denotes a group of the cross terms and:

$$\begin{aligned} \widehat{R}_{T,\widehat{\varepsilon},1}(h) &= \sum_{t=1}^{T} \left\{ \widehat{f}(\varepsilon_t) + \tilde{f}(\varepsilon_t, \tilde{\theta}) \right\}^2 + \sum_{t=1}^{T} \left\{ \widehat{f}(\widehat{\varepsilon}_t) - \widehat{f}(\varepsilon_t) \right\}^2 + \sum_{t=1}^{T} \left\{ \tilde{f}(\widehat{\varepsilon}_t, \tilde{\theta}) - \tilde{f}(\varepsilon_t, \tilde{\theta}) \right\}^2 \\ &= \widehat{R}_{T,\varepsilon}(h) + \widehat{r}_{T,1}(h) + \widehat{r}_{T,2}(h). \end{aligned}$$

Note that $\widehat{r}_{T,1}(h)$ can be dealt with in a fashion similar to the proof of Lemma C.4 of Saart et al. (2013). With regard to $\widehat{r}_{T,2}(h)$, we have

$$\begin{aligned} &\tilde{f}(\widehat{\varepsilon}_t, \tilde{\theta}) = \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\widehat{\varepsilon}_s, \tilde{\theta}) = \sum_{s=1}^{T} w_s(\varepsilon_t) f(\varepsilon_s, \tilde{\theta}) + \left\{ \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\widehat{\varepsilon}_s, \tilde{\theta}) - \sum_{s=1}^{T} w_s(\varepsilon_t) f(\varepsilon_s, \tilde{\theta}) \right\} \\ &= \tilde{f}(\varepsilon_t, \tilde{\theta}) + \left\{ \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\widehat{\varepsilon}_s, \tilde{\theta}) - \sum_{s=1}^{T} w_s(\varepsilon_t) f(\varepsilon_s, \tilde{\theta}) \right\} = \tilde{f}(\varepsilon_t, \tilde{\theta}) + \Delta_{\tilde{f}}, \end{aligned}$$

where

$$\Delta_{\tilde{f}} = \left\{ \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\widehat{\varepsilon}_s, \tilde{\theta}) - \sum_{s=1}^{T} w_s(\varepsilon_t) f(\varepsilon_s, \tilde{\theta}) \right\} = \Delta_{\tilde{f},1} + \Delta_{\tilde{f},2} \tag{A.21}$$

by which

$$\Delta_{\tilde{f},1} = \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\varepsilon_s, \tilde{\theta}) - \sum_{s=1}^{T} w_s(\varepsilon_t) f(\varepsilon_s, \tilde{\theta}) = \sum_{s=1}^{T} \{ w_s(\widehat{\varepsilon}_t) - w_s(\varepsilon_t) \} f(\varepsilon_s, \tilde{\theta}) \tag{A.22}$$

and

$$\Delta_{\tilde{f},2} = \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\varepsilon_s, \tilde{\theta}) - \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) f(\widehat{\varepsilon}_s, \tilde{\theta}) = \sum_{s=1}^{T} w_s(\widehat{\varepsilon}_t) \left\{ f(\varepsilon_s, \tilde{\theta}) - f(\widehat{\varepsilon}_s, \tilde{\theta}) \right\}. \tag{A.23}$$

We also have

$$\begin{aligned} &w_s(\widehat{\varepsilon}_t) - w_s(\varepsilon_t) = \{ w_s(\widehat{\varepsilon}_t, h) - w_s(\varepsilon_t, h) \} \\ &= \left\{ \frac{1}{T} k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) \left[\frac{(s_2(\widehat{\varepsilon}_t) - s_1(\widehat{\varepsilon}_t)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s))}{(s_2(\widehat{\varepsilon}_t) s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t))} \right] \right\} - \left\{ \frac{1}{T} k_h(\varepsilon_t - \varepsilon_s) \left[\frac{(s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s))}{(s_2(\varepsilon_t) s_0(\varepsilon_t) - s_1^2(\varepsilon_t))} \right] \right\} \\ &= \Delta_{w_{s,1}} + \Delta_{w_{s,2}}, \end{aligned}$$

where

$$\Delta_{w_{s,1}} = \frac{1}{T}k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\left\{\left[\frac{(s_2(\widehat{\varepsilon}_t) - s_1(\widehat{\varepsilon}_t)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s))}{(s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t))}\right] - \left[\frac{(s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s))}{(s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t))}\right]\right\}$$

and

$$\Delta_{w_{s,2}} = \left[\frac{(s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s))}{(s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t))}\right]\left\{\frac{1}{T}k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - \frac{1}{T}k_h(\varepsilon_t - \varepsilon_s)\right\}.$$

Note that

$$\left[\frac{(s_2(\widehat{\varepsilon}_t) - s_1(\widehat{\varepsilon}_t)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s))}{(s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t))}\right] - \left[\frac{(s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s))}{(s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t))}\right] = \Delta_{s,1} - \Delta_{s,2},$$

where

$$\Delta_{s,1} = \frac{\{s_2(\widehat{\varepsilon}_t) - s_1(\widehat{\varepsilon}_t)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\} - \{s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s)\}}{(s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t))} \quad \text{and}$$

$$\Delta_{s,2} = \left\{\frac{s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s)}{s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t)}\right\}\left\{\frac{\{s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t)\} - \{s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t)\}}{s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t)}\right\}.$$

Therefore, $\Delta_{\widetilde{f},1}$ can be bounded by the following quantity:

$$\begin{aligned}\frac{1}{T}\left|\sum_{s=1}^{T}\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\}\{\Delta_{s,1} - \Delta_{s,2}\}f(\varepsilon_s, \widetilde{\theta})\right| &\leq \frac{1}{T}\left|\sum_{s=1}^{T}\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\}\Delta_{s,1}f(\varepsilon_s, \widetilde{\theta})\right| \\ &+ \frac{1}{T}\left|\sum_{s=1}^{T}\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\}\Delta_{s,2}f(\varepsilon_s, \widetilde{\theta})\right|. \qquad \text{(A.24)}\end{aligned}$$

Let us now concentrate on $\Delta_{s,j}$ for $j = 1, 2$. Observe that $\Delta_{s,1} = \frac{\Delta_{s,11}}{\Delta_{s,12}}$, where

$$\begin{aligned}|\Delta_{s,11}| &= |\{s_2(\widehat{\varepsilon}_t) - s_1(\widehat{\varepsilon}_t)(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\} - \{s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s)\}| \\ &\leq |s_1(\widehat{\varepsilon}_t) - s_1(\varepsilon_t)| + |s_1(\varepsilon_t)|\{|\widehat{\varepsilon}_t - \varepsilon_t| + |\widehat{\varepsilon}_s - \varepsilon_s|\} \leq o_P\{T^{-\delta}\}, \qquad \text{(A.25)}\end{aligned}$$

$$\Delta_{s,12} = s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t) = s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t) + \Delta_{s,12R} \qquad \text{(A.26)}$$

and

$$\begin{aligned}|\Delta_{s,12R}| &\leq |s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_2(\varepsilon_t)s_0(\varepsilon_t)| + |s_1(\widehat{\varepsilon}_t) - s_1(\varepsilon_t)|^2 + 2|s_1(\varepsilon_t)||s_1(\widehat{\varepsilon}_t) - s_1(\varepsilon_t)| \\ &\leq o_P\{T^{-\delta}\} \qquad \text{(A.27)}\end{aligned}$$

following Lemma A.2 and Assumption 2.1. With regard to $\Delta_{s,2}$, observe also that

$$\begin{aligned}|\Delta_{s,2}| &= \left|\left\{\frac{s_2(\varepsilon_t) - s_1(\varepsilon_t)(\varepsilon_t - \varepsilon_s)}{s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t)}\right\}\left\{\frac{1}{\{s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t)\} + o_P\{1\}}\right\}\Delta_{s,2R}\right| \\ &\leq o_P\{T^{-\delta}\}, \qquad \text{(A.28)}\end{aligned}$$

where $\Delta_{s,2R} = \left\{\{s_2(\widehat{\varepsilon}_t)s_0(\widehat{\varepsilon}_t) - s_1^2(\widehat{\varepsilon}_t)\} - \{s_2(\varepsilon_t)s_0(\varepsilon_t) - s_1^2(\varepsilon_t)\}\right\}$, following a similar set of arguments as in (A.25) and (A.27). Therefore, in (A.24) we have for $j = 1, 2$,

$$\begin{aligned}\frac{1}{T}\sum_{s=1}^{T}k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s)\Delta_{s,j}f(\varepsilon_s, \widetilde{\theta}) &= \frac{1}{T}\sum_{s=1}^{T}\{k_h(\widehat{\varepsilon}_t - \widehat{\varepsilon}_s) - k_h(\varepsilon_t - \varepsilon_s)\}\Delta_{s,j}f(\varepsilon_s, \widetilde{\theta}) \\ &+ \frac{1}{T}\sum_{s=1}^{T}k_h(\varepsilon_t - \varepsilon_s)\Delta_{s,j}f(\varepsilon_s, \widetilde{\theta}), \qquad \text{(A.29)}\end{aligned}$$

where we have

$$\left|\frac{1}{T}\sum_{s=1}^{T}\left\{k_h(\widehat{\varepsilon}_t-\widehat{\varepsilon}_s)-k_h(\varepsilon_t-\varepsilon_s)\right\}\Delta_{s,j}f(\varepsilon_s,\widetilde{\theta})\right|$$

$$\leq \sup_{h\in H_T}\max_{1\leq t\leq T}|\Delta_{s,j}|\left\{\frac{1}{T}\sum_{s=1}^{T}|k_h(\widehat{\varepsilon}_t-\widehat{\varepsilon}_s)-k_h(\varepsilon_t-\varepsilon_s)|\right\}\leq o_P\{T^{-\delta}\} \quad \text{and} \tag{A.30}$$

$$\left|\frac{1}{T}\sum_{s=1}^{T}k_h(\varepsilon_t-\varepsilon_s)\Delta_{s,j}f(\varepsilon_s,\widetilde{\theta})\right|\leq \sup_{h\in H_T}\max_{1\leq t\leq T}|\Delta_{s,j}|\left\{\frac{1}{T}\sum_{s=1}^{T}|k_h(\varepsilon_t-\varepsilon_s)|\right\}\leq o_P\{T^{-\delta}\} \tag{A.31}$$

following (A.8) to (A.10) and (A.7). Finally, the required results can be obtained for each of the term included in $\widehat{R}_{T,\widehat{\varepsilon},2}(h)$ using Cauchy–Schwarz inequality and the convergence rate of $\widehat{r}_{T,1}(h)$ and $\widehat{r}_{T,2}(h)$. ■

Lemma A.4. *Let Assumptions A.1 and A.2 hold. In addition, let*

$$\widehat{S}_{T,\widehat{\varepsilon}}(h)=\frac{1}{T}\sum_{t=1}^{T}\widehat{f}(\widehat{\varepsilon}_t) \text{ and } \widehat{S}_{T,\varepsilon}(h)=\frac{1}{T}\sum_{t=1}^{T}\widehat{f}(\varepsilon_t).$$

Then, we have, as $T\to\infty$,

$$\widehat{S}_{T,\widehat{\varepsilon}}(h)=\widehat{S}_{T,\varepsilon}(h)+o_P\{1\} \quad \text{uniformly over } h\in H_T. \tag{A.32}$$

Proof. Observe that

$$\widehat{S}_{T,\widehat{\varepsilon}}(h)-\widehat{S}_{T,\varepsilon}(h)=\frac{1}{T}\sum_{t=1}^{T}\widehat{f}(\widehat{\varepsilon}_t)-\frac{1}{T}\sum_{t=1}^{T}\widehat{f}(\varepsilon_t)$$

$$=\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{T}\sum_{s=1}^{T}k_h(\widehat{\varepsilon}_t-\widehat{\varepsilon}_s)-\frac{1}{T}\sum_{s=1}^{T}k_h(\varepsilon_t-\varepsilon_s)\right\}=\frac{1}{T}\sum_{t=1}^{T}\Delta_{k_h}(h). \tag{A.33}$$

The required result is obtained immediately by (A.11). ■

Lemma A.5. *Let Assumptions A.1 and A.2 hold. In addition, let*

$$\widehat{U}_{T,\widehat{\varepsilon}}(h)=\frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{f}(\widehat{\varepsilon}_t)\right\}^3 \text{ and } \widehat{U}_{T,\varepsilon}(h)=\frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{f}(\varepsilon_t)\right\}^3.$$

Then, we have, as $T\to\infty$,

$$\widehat{U}_{T,\widehat{\varepsilon}}(h)=\widehat{U}_{T,\varepsilon}(h)+o_P\{1\} \text{ uniformly over } h\in H_T. \tag{A.34}$$

Proof. Observe that for some $0<C<\infty$

$$\begin{aligned}\left|\widehat{U}_{T,\widehat{\varepsilon}}(h)-\widehat{U}_{T,\varepsilon}(h)\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{f}(\widehat{\varepsilon}_t)\right\}^3-\frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{f}(\varepsilon_t)\right\}^3\right|\leq\frac{1}{T}\sum_{t=1}^{T}\left|\left(\left\{\frac{\widehat{f}(\widehat{\varepsilon}_t)}{\widehat{f}(\varepsilon_t)}\widehat{f}(\varepsilon_t)\right\}^3-\left\{\widehat{f}(\varepsilon_t)\right\}^3\right)\right| \\ &= \frac{1}{T}\sum_{t=1}^{T}\left|\{f(\varepsilon_t)\}^3\left(\left\{\frac{\widehat{f}(\widehat{\varepsilon}_t)}{\widehat{f}(\varepsilon_t)}\right\}^3-1\right)\right|\left\{\frac{\widehat{f}(\varepsilon_t)}{f(\varepsilon_t)}\right\}^3 \\ &\leq C\sup_{h\in H_T}\max_{1\leq t\leq T}\left|\left(\frac{\widehat{f}(\widehat{\varepsilon}_t)}{\widehat{f}(\varepsilon_t)}\right)^3-1\right|\cdot\int f^4(\varepsilon)d\varepsilon. \end{aligned} \tag{A.35}$$

Then the required result is obtained immediately using the convergence result in (A.11). ■

Lemma A.6. *Let Assumptions A.1 and A.2 hold. Then we have, as* $T \to \infty$,

$$\text{(i) } \widehat{L}_{T,\varepsilon}(h) = L_{T,\varepsilon}(h) + o_P(1) \text{ and (ii) } L_{T,\varepsilon}(h) = \widetilde{L}_{T,\varepsilon}(h) + o_P(1)$$

uniformly in $h \in H_T$, *where* $L_{T,\varepsilon}(h) = \frac{\widehat{N}_{T,\varepsilon}(h)-\mu_0}{\sqrt{h}\sigma_0}$ *and* $\widetilde{L}_{T,\varepsilon}(h) = \frac{\widetilde{N}_{T,\varepsilon}(h)-\mu_0}{\sqrt{h}\sigma_0}$, *in which* $\widetilde{N}_{T,\varepsilon}(h) = Th \cdot \int \left(\widehat{f}(x) - \tilde{f}(x,\tilde{\theta})\right)^2 \widehat{f}(x)dx$.

Proof. The proof of Lemma A.6(i) follows from that of lemma B.6 of Gao and King (2004). Hence, we start by verifying Lemma A.6(ii). Observe that by the uniform convergence of $\widehat{f}(\cdot)$ to $f(\cdot)$ as well as the weak law of large numbers (as used in the proof of theorem 2.2 of Gao and King (2004)), as $T \to \infty$, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{f}(\varepsilon_t) - \tilde{f}(\varepsilon_t,\theta)\right)^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{f}(\varepsilon_t) - f(\varepsilon_t,\theta)\right)^2 + o_P(1) \qquad \text{(A.36)}\\
&= \frac{1}{T}\sum_{t=1}^{T}\left(f(\varepsilon_t) - f(\varepsilon_t,\theta)\right)^2 + o_P(1) = \int_{-\infty}^{\infty}\left(f(x) - f(x,\theta)\right)^2 f(x)dx + o_P(1),\\
\int_{-\infty}^{\infty}\left(\widehat{f}(x) - \tilde{f}(x,\theta)\right)^2 \widehat{f}(x)dx &= \int_{-\infty}^{\infty}\left(\widehat{f}(x) - f(x,\theta)\right)^2 \widehat{f}(x)dx + o_P(1)\\
&= \int_{-\infty}^{\infty}\left(f(x) - f(x,\theta)\right)^2 f(x)dx + o_P(1). \qquad \text{(A.37)}
\end{aligned}$$

The verification of Lemma A.6(ii) then follows from (A.36) and (A.37).

A.3. Proofs of the Main Results in Section 2:

Proof of Lemma 2.1: We need to show that:

$$\begin{aligned}
\widehat{L}_{T,\widehat{\varepsilon}}(h) &= \frac{\widehat{N}_{T,\widehat{\varepsilon}}(h) - \widehat{\mu}_{T,\widehat{\varepsilon}}(h)}{\sqrt{h}\widehat{\sigma}_{T,\widehat{\varepsilon}}(h)}\\
&= \frac{\{\widehat{N}_{T,\varepsilon}(h) - \widehat{\mu}_{T,\varepsilon}(h)\} + \left(\{\widehat{N}_{T,\widehat{\varepsilon}}(h) - \widehat{N}_{T,\varepsilon}(h)\} - \{\widehat{\mu}_{T,\widehat{\varepsilon}}(h) - \widehat{\mu}_{T,\varepsilon}(h)\}\right)}{\sqrt{h}\widehat{\sigma}_{T,\varepsilon}(h) + \sqrt{h}\left(\widehat{\sigma}_{T,\widehat{\varepsilon}}(h) - \widehat{\sigma}_{T,\varepsilon}(h)\right)}\\
&= \frac{\{\widehat{N}_{T,\varepsilon}(h) - \widehat{\mu}_{T,\varepsilon}(h)\}}{\sqrt{h}\widehat{\sigma}_{T,\varepsilon}(h)} + o_P\{T^{-\delta}\}, \qquad \text{(A.38)}
\end{aligned}$$

where

$$\begin{aligned}
\widehat{N}_{T,\widehat{\varepsilon}}(h) - \widehat{N}_{T,\varepsilon}(h) &= h \cdot \left\{\sum_{t=1}^{T}(\widehat{f}(\widehat{\varepsilon}_t) - \tilde{f}(\widehat{\varepsilon}_t,\tilde{\theta}))^2 - \sum_{t=1}^{T}(\widehat{f}(\varepsilon_t) - \tilde{f}(\varepsilon_t,\tilde{\theta}))^2\right\}\\
&= h \cdot \left\{\widehat{R}_{T,\widehat{\varepsilon}}(h) - \widehat{R}_{T,\varepsilon}(h)\right\},\\
\widehat{\mu}_{T,\widehat{\varepsilon}}(h) - \widehat{\mu}_{T,\varepsilon}(h) &= R(k) \cdot \left\{\frac{1}{T}\sum_{t=1}^{T}\widehat{f}(\widehat{\varepsilon}_t) - \frac{1}{T}\sum_{t=1}^{T}\widehat{f}(\varepsilon_t)\right\}\\
&= R(k) \cdot \left\{\widehat{S}_{T,\widehat{\varepsilon}}(h) - \widehat{S}_{T,\varepsilon}(h)\right\},\\
\widehat{\sigma}^2_{T,\widehat{\varepsilon}}(h) - \widehat{\sigma}^2_{T,\varepsilon}(h) &= 2k^{(4)}(0) \cdot \left\{\frac{1}{T}\sum_{t=1}^{T}\{\widehat{f}(\widehat{\varepsilon}_t)\}^3 - \frac{1}{T}\sum_{t=1}^{T}\{\widehat{f}(\varepsilon_t)\}^3\right\}\\
&= 2k^{(4)}(0) \cdot \left\{\widehat{U}_{T,\widehat{\varepsilon}}(h) - \widehat{U}_{T,\varepsilon}(h)\right\}. \qquad \text{(A.39)}
\end{aligned}$$

The required results are obtained immediately following Lemmas A.3 to A.5 above. ■

Proof of Theorem 2.1: In view of Lemmas 2.1 and A.6, the proof of Theorem 2.1 follows immediately from Theorem 2.2 of Gao and King (2004). ■

Proof of Theorem 2.2: In view of Lemmas 2.1 and A.6, the proof of Theorem 2.2 follows immediately from Lemma B.7 of Gao and King (2004). ■

Proof of Theorems 2.3 and 2.4: In view of Lemmas 2.1 and A.6, the proof of Theorems 2.3 and 2.4 follows immediately from Lemmas B.1 and B.7 of Gao and King (2004). ■

8. Appendix B

In this Appendix we first provide a brief review of the SEMI-ACD model and its estimation procedure. A more detailed discussion can be found in Saart et al. (2013). We then presents the proof of the main results of Section 3.1.

It should be noted that in this section the bandwidth parameter, h, is associated with the estimation of the SEMI-ACD model, unlike that in the previous sections, which was the bandwidth parameter for the computation of the test statistic. However, since the material presented here is self-contained, we also denote this bandwidth parameter by h.

B.1. SEMI-ACD Model

The SEMI-ACD(1,1) model proposed by Saart et al. (2013) defines

$$\psi_t = \gamma x_{t-1} + g\left(\psi_{t-1}\right), \tag{B.1}$$

where γ is an unknown parameter and $g(\cdot)$ is an unknown function satisfying the Lipschitz-type condition in Assumption B.1.

The estimation of the model is based on an algorithmically computed estimate of the t-th conditional duration at the m^*th iteration defined by

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

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

$$\widehat{g}_{j,h}(\widehat{\psi}_{t-1,m^*-1}) = \sum_{s=m^*+\iota}^{T} W_{s,h}(\widehat{\psi}_{t-1,m^*-1})x_{s-j+1} \tag{B.3}$$

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

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

where $k_h(\cdot) = h^{-1}k(\cdot/h)$, k is a real-valued kernel function satisfying Assumption B.3 and the bandwidth $h = h_T \in H_T = \left[a_1 T^{-\frac{1}{5}-c_1}, b_1 T^{-\frac{1}{5}+c_1}\right]$, in which $0 < a_1 < b_1 < \infty$ and $0 < c_1 < \frac{1}{20}$.

Furthermore, the kernel weighted least squares estimators of γ and $\sigma^2 = E(\eta_1^2)$ are written as

$$\widehat{\gamma}_{\widehat{\psi}}(h) = \left\{\sum_{t=1}^{T} \widehat{u}_{t+1}^2 \omega(\widehat{\psi}_{t,m^*})\right\}^{-1} \left\{\sum_{t=1}^{T} \widehat{u}_{t+1}\left(x_{t+1} - \widehat{g}_{1,h}(\widehat{\psi}_{t,m^*})\right)\omega(\widehat{\psi}_{t,m^*})\right\} \tag{B.4}$$

and

$$\widehat{\sigma}^2_{\widehat{\psi}}(h) = \frac{1}{T}\sum_{t=1}^{T}\{x_{t+1} - \widehat{\gamma}_{\widehat{\psi}}(h)x_t - \widehat{g}_{1,h}(\widehat{\psi}_{t,m^*}) + \widehat{\gamma}_{\widehat{\psi}}(h)\widehat{g}_{2,h}(\widehat{\psi}_{t,m^*})\}^2\omega(\widehat{\psi}_{t,m^*}), \tag{B.5}$$

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.2(ii).

In order to proceed with the hypothesis testing introduced in this paper, we introduce the following algorithm-based estimate of the standardized duration, i.e. the SEMI–ACD residual

$$\widehat{\varepsilon}_{t,m^*} = \frac{x_t}{\widehat{\psi}_{t,m^*}}, \tag{B.6}$$

where $\widehat{\psi}_{t,m^*}$ is as defined in (B.2). Now, observe that we have

$$|\widehat{\varepsilon}_{t,m^*} - \varepsilon_t| = \left| \frac{x_t}{\widehat{\psi}_{t,m^*}} - \frac{x_t}{\psi_t} \right| = \left\{ \frac{\varepsilon_t}{\psi_t} \right\} \left| \widehat{\psi}_{t,m^*} - \psi_t \right| \left\{ \frac{\psi_t}{\widehat{\psi}_{t,m^*}} \right\}. \tag{B.7}$$

Hence, a uniform consistency of $\widehat{\psi}_{t,m^*}$, for example, immediately leads to a similar mode of consistency of $\widehat{\varepsilon}_{t,m^*}$. Additional discussion of the iterative estimation algorithm used can be found in Saart et al. (2013).

B.2. Technical Assumptions

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

$$|g(y) - g(x)| \le \varphi(x)|y - x| \tag{B.8}$$

for each given x and $y \in S_\omega$, where S_ω is the compact support of the weight function $\omega(\cdot)$ 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) Let $z_t = x_{t-1} - g_2(\psi_{t-1})$ 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 $\delta_2 > 0$.*

(ii) Suppose that the nonnegative weight function $\omega(\cdot)$ is continuous and bounded.

Assumption B.3. *Let Assumption 2.2 holds.*

B.3. Proofs of Lemmas 3.1 and 3.2

Proof of Lemma 3.1 Observe firstly that for any $m \ge 1$ and $t \ge m$, we have:

$$\begin{aligned} \left|\widehat{\psi}_{t,m} - \psi_t\right| &\le |\widehat{\gamma}_m(h) - \gamma|\,|x_{t-1}| + \left|g(\widehat{\psi}_{t-1,m-1}) - g(\psi_{t-1})\right| + \left|\widehat{g}^*_h(\widehat{\psi}_{t-1,m-1}) - g(\widehat{\psi}_{t-1,m-1})\right| \\ &\le |\widehat{\gamma}_m(h) - \gamma| \cdot \left(\sum_{j=2}^{m}\prod_{d=2}^{j}\varphi(\psi_{t-d+1})\,|x_{t-j}|\right) + \prod_{j=2}^{m}\varphi(\psi_{n-j+1})\left|\widehat{\psi}_{t-m,0} - \psi_{t-m}\right| \\ &\quad + o_P\left(N^{-\frac{1}{4}}\right) \cdot \left(\sum_{j=2}^{m}\prod_{d=2}^{j}\varphi(\psi_{t-d+1})\right), \end{aligned} \tag{B.9}$$

where the second inequality holds by iteration, the uniform consistency of $\widehat{g}_{j,h}(\cdot)$ for $j = 1, 2$ (see Lemma C.3 of Saart et al. (2013)) and $\sup_{x\in S_\omega} |g_j(x)| \le B_g < \infty$. In view of (B.9), Assumption B.1, the facts that $E\left[\varphi(\psi_{n+1})|(\psi_n, \cdots, \psi_1)\right] \le E^{\frac{1}{2}}\left[\varphi^2(\psi_{n+1})|(\psi_n, \cdots, \psi_1)\right] \le G$ with probability one and that $E[\psi_1^2] < \infty$ and $E[x_1^2] < \infty$, suggest that we have by choosing m^* as recommended in their Lemma C.3

$$\left|\widehat{\psi}_{t,m^*} - \psi_t\right| \leq \sup_{h\in H_T} \max_{t\geq m+1} \left|\widehat{\psi}_{t,m^*} - \psi_t\right| = o_P\left\{T^{-\frac{1}{4}}\right\}. \tag{B.10}$$

Observe also that

$$\begin{aligned} |\widehat{\varepsilon}_{t,m^*} - \varepsilon_t| = \left|\frac{x_t}{\widehat{\psi}_{t,m^*}} - \frac{x_t}{\psi_t}\right| &= \left|\left\{\frac{\varepsilon_t}{\psi_t}\right\}\left\{\widehat{\psi}_{t,m^*} - \psi_t\right\}\left\{\frac{\psi_t}{\widehat{\psi}_{t,m^*}}\right\}\right| \leq \sup_{h\in H_T} \max_{t\geq m+1} \left\{\frac{x_t}{\psi_t^2}\right\}\left|\widehat{\psi}_{t,m} - \psi_t\right| \\ &= o_P\left\{T^{-\frac{1}{4}}\right\} \end{aligned} \tag{B.11}$$

as claimed by (3.6) and because $\left\{\frac{\psi_t}{\widehat{\psi}_{t,m^*}}\right\} = 1 + o_P\{1\}$. ■

Proof of Lemma 3.2 We obtain by recursion

$$\begin{aligned} \widehat{\psi}_{t,Qml} - \psi_t &= \left\{(\widehat{\alpha} - \alpha)x_{t-1} - (\widehat{\beta} - \beta)\psi_{t-1,Qml}\right\} \\ &- \widehat{\beta}\left\{(\widehat{\alpha} - \alpha)x_{t-2} - (\widehat{\beta} - \beta)\psi_{t-2,Qml}\right\} - \cdots - \widehat{\beta}^T\left\{(\widehat{\alpha} - \alpha)x_{t-T} - (\widehat{\beta} - \beta)\psi_{t-T,Qml}\right\} \\ &- \widehat{\beta}^T\left\{\widehat{\psi}_{T,Qml} - \psi_T\right\}. \end{aligned}$$

Furthermore, note that $|\beta| < 1$ due to the stationarity condition. The $\sqrt{n}$-consistency of the QML estimators shows that we have $\widehat{\alpha} - \alpha = o_P\left\{T^{-\frac{1}{2}}\right\}$ and, similarly, $\widehat{\beta} - \beta = o_P\left\{T^{-\frac{1}{2}}\right\}$. Finally, the moment conditions in Assumption B.2 imply immediately that

$$\left|\widehat{\psi}_{t,Qml} - \psi_t\right| \leq \max_{1\leq t\leq T} \left|\widehat{\psi}_{t,Qml} - \psi_t\right| = o_P\{T^{-1/2}\}. \tag{B.12}$$

The required result for $\widehat{\varepsilon}_{t,Qml}$ can then be obtained as in (B.11) above. ■

B.4. Proof of Corollaries 3.1 to 3.3

The proofs of Corollaries 3.1, 3.2 and 3.3 follow closely from those of Theorems 2.2, 2.3 and 2.4. Nonetheless, along the way we replace the estimator in Assumption 2.1 with $\widehat{\varepsilon}_{t,m^*}$ and $\widehat{\varepsilon}_{t,Qml}$, and use their corresponding uniform convergence as stated in Lemmas 3.1 and 3.2, respectively. ■

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.

Bowsher, C. G., 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. Journal of Econometrics 141 (2), 876–912.

Broock, W., Scheinkman, J. A., Dechert, W. D., LeBaron, B., 1996. A test for independence based on the correlation dimension. Econometric Reviews 15 (3), 197–235.

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

Chou, R. Y., 2005. Forecasting financial volatilities with extreme values: the conditional autoregressive range (carr) model. Journal of Money, Credit and Banking, 561–582.

Diebold, F. X., Gunther, T., Tsay, A., 1998. Evaluating density forecasts with applications to financial risk managment. International Economic Review 39, 863–883.

Dufour, A., Engle, R. F., 2000. Time and the price impact of a trade. The Journal of Finance 55 (6), 2467–2498.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, 987–1007.

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

Engle, R. F., 2002. New frontiers for arch models. Journal of Applied Econometrics 17 (5), 425–446.

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

Fan, Y., 1994. Testing the goodness of fit of a parametric density function by kernel method. Econometric Theory 10, 316–316.

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

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

Gao, J., Gijbels, I., 2008. Bandwidth selection in nonparametric kernel testing. Journal of the American Statistical Association 103 (484), 1584–1594.

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

Ghysels, E., Gourieroux, C., Jasiak, J., 2004. Stochastic volatility duration models. Journal of Econometrics 119 (2), 413–433.

Giesecke, K., Kim, B., 2011. Systemic risk: What defaults are telling us. Management Science 57 (8), 1387–1405.

Giesecke, K., Tomecek, P., 2005. Dependent events and changes of time. Permanent Working Paper.

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.

Hong, Y., Lee, T.-H., 2003. Diagnostic checking for the adequacy of nonlinear time series models. Econometric theory 19 (6), 1065–1121.

Hong, Y., Lee, Y.-J., 2011. Detecting misspecifications in autoregressive conditional duration models and non-negative time-series processes. Journal of Time Series Analysis 32 (1), 1–32.

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

Li, W. K., Yu, P. L. H., 2003. On the residual autocorrelation of the autoregressive conditional duration model. Economics Letters 79 (2), 169–175.

Mammen, E., Rothe, C., Schienle, M., 2012. Nonparametric regression with nonparametrically generated covariates. The Annals of Statistics 40 (2), 1132–1170.

Manganelli, S., 2005. Duration, volume and volatility impact of trades. Journal of Financial markets 8 (4), 377–399.

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

Pacurar, M., 2008. Autoregressive conditional duration models in finance: a survey of the theoretical and empirical literature. Journal of Economic Surveys 22 (4), 711–751.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. The annals of mathematical statistics 23 (3), 470–472.

Saart, P. W., Gao, J., Allen, D., 2013. Semiparametric autoregressive conditional duration model: Theory and practice. Econometric Reviews (In press) (working paper version with detailed proof is available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2109200`).

Sperlich, S., 2009. A note on non-parametric estimation with predicted variables. The Econometrics Journal 12 (2), 382–395.

Taylor, S. J., 1982. Financial returns modelled by the product of two stochastic rpocesses – a study of daily sugar prices. In: Anderson, O. D. (Ed.), Time series analysis: theory and practice. North-Holland.

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