Original Article

Journal of
Time Series Analysis

First version received 5 September 2008

Published online in Wiley Online Library: 2 September 2010

(wileyonlinelibrary.com) DOI: 10.1111/j.1467-9892.2010.00681.x

Detecting misspecifications in autoregressive conditional duration models and non-negative time-series processes

Yongmiao Hong[a,*,†] and Yoon-Jin Lee[b,*,†]

We develop a general theory to test correct specification of multiplicative error models of non-negative time-series processes, which include the popular autoregressive conditional duration (ACD) models. Both linear and nonlinear conditional expectation models are covered, and standardized innovations can have time-varying conditional dispersion and higher-order conditional moments of unknown form. No specific estimation method is required, and the tests have a convenient null asymptotic N(0,1) distribution. To reduce the impact of parameter estimation uncertainty in finite samples, we adopt Wooldridge's (1990a) device to our context and justify its validity. Simulation studies show that in the context of testing ACD models, finite sample correction gives better sizes in finite samples and are robust to parameter estimation uncertainty. And, it is important to take into account time-varying conditional dispersion and higher-order conditional moments in standardized innovations; failure to do so can cause strong overrejection of a correctly specified ACD model. The proposed tests have reasonable power against a variety of popular linear and nonlinear ACD alternatives.

Keywords: Autoregressive conditional duration; dispersion clustering; finite sample correction; generalized spectral derivative; nonlinear time series; parameter estimation uncertainty; Wooldridge's Device
JEL classifications: C4; C2.

1. INTRODUCTION

High-frequency data have become widely available in economics and finance over the past decade. As a result of the availability of these data sets and the rapid advance in computing power, there is a growing interest in modelling high-frequency financial data. The analysis of high-frequency data has rapidly developed as a promising research area by facilitating a deeper understanding of market activity. As Engle and Russell (1998) point out, quantity purchased in a period of time is often the key economic variable to be modelled or forecast, and market microstructure theories are typically tested on a transaction-by-transaction basis. Such massive transaction data provide rich information about financial activities and market microstructure.

In high-frequency financial econometrics, the timing of transactions is a key factor to understanding economic theory. For example, the time duration between market events has been found to have a deep impact on the behaviour of market agents (e.g. traders and market makers) and on the intraday characteristics of the price process. Recent models in market microstructure literature based on asymmetric information argue that time may convey information and should be modelled as well. The important role of time has been highlighted by Easley and O'Hara (1992) and Easley *et al.* (1997), which generalize Glosten and Milgrom (1985).

However, an inherent feature of transaction data presents a great challenge to econometricians. When every single transaction and quoted price are recorded, the ultimate limit case, 'ultra-high frequency data' as termed by Engle (2000), is obtained. Consequently, the arrival times of events (e.g. quotes, trades) are irregularly spaced, and the time between successive observations is not deterministic but random. This renders standard time-series econometric tools inapplicable, since they are based on fixed, regularly spaced time interval analysis. Motivated by this feature, Engle and Russell (1998) and Engle (2000) propose a class of autoregressive conditional duration (ACD) models to characterize the arrival time intervals between market events of interest such as the occurrence of a trade or a bid-ask quote. The main idea behind ACD modelling is a dynamic parameterization of the conditional expected duration given the past information. This model combines elements of time-series models and econometric tools for analysing transition data (e.g. Lancaster, 1990) and is well suited for the analysis of high-frequency financial data. In addition, ACD models are used as a building block for jointly modelling duration and other market characteristics (e.g. price and volume), which may improve understanding of the complex nature of a trading process.

[a]*Cornell University and Xiamen University*
[b]*Indiana University*
**Correspondence to: Yongmiao Hong, Department of Economics and Department of Statistical Science, Uris Hall, Cornell University, Ithaca, NY 14850, USA, and Wang Yanan Institute for Studies in Economics (WISE), and MOE Key Laboratory in Econometrics, Xiamen University, Xiamen 361005, China; Yoon-Jin Lee, Department of Economics, Indiana University, Wylie Hall, Bloomington, IN 47405, USA.*
[†]*E-mail: yh20@cornell.edu; lee243@indiana.edu*

Despite the vast literature on ACD specifications, model evaluation has not yet received much attention, as pointed out in (e.g.) Li and Yu (2003), Meitz and Teräsvirta (2006) and Pacurar (2006). In particular, formal evaluation of ACD models via specification testing has not been common in empirical study. Most works limit the testing to simple examinations of the standardized residual. Since the flexibility of ACD models arises from various choices of the models for the conditional expected duration and the probability density of the standardized innovations, there have been two categories of specification tests for ACD models. The first checks the probability distribution specification for standardized innovations. Bauwens *et al.* (2004) check the goodness-of-fit of an ACD model using the density forecast evaluation methods of Diebold *et al.* (1998). Fernandes and Grammig (2005) consider nonparametric specification tests against distributional misspecifications of the standardized innovations, assuming that the conditional expected duration model is correctly specified. The second category of tests checks specification for conditional expected duration. The Box–Pierce–Ljung type portmanteau test statistic is often applied to the estimated standardized or squared estimated standardized durations, as in (e.g.) Dufour and Engle (2000), Bauwen *et al.* (2004) and Fernandes and Grammig (2005). However, the Box–Pierce–Ljung type test, when applied to estimated standardized durations, is invalid even asymptotically, because it does not take into account the impact of parameter estimation uncertainty on the asymptotic distribution of the test statistic. Li and Yu (2003) derive an asymptotically valid modified portmanteau test for an ACD model based on the estimated standardized duration autocorrelations in the spirit similar to Li and Mak (1994). Meitz and Teräsvirta (2006) develop a class of Lagrange multiplier (LM) tests for an ACD model against various parametric alternatives for conditional expected duration. One of their tests is asymptotically equivalent to Li and Yu's (2003) test. Hautsch (2006) also considers some LM tests as well as various conditional moment tests and generalized conditional moment tests for conditional expected duration specification. Building on Hong (1996, 1997), Duchesne and Pacurar (2006) construct tests for the adequacy of ACD models, based on a kernel spectral density estimator of the standardized innovation process. They obtain a generalized version of the classical Box–Pierce–Ljung test statistic as a special case. In the literature, the i.i.d. test of Hong and Lee (2003) has been also used to test conditional expected duration (e.g. Meitz and Teräsvirta, 2006). This is not suitable when standardized innovations are not i.i.d. because it would reject a correctly specified conditional expected duration model when standardized innovations display dependence in higher-order moments (e.g. time-varying dispersion).

In this article, we propose a new class of specification tests for the conditional expected duration. Specification for the conditional expected duration is a fundamental building block of an ACD model. Correct specification of conditional expected duration dynamics is required to ensure consistency of the quasi-maximum likelihood estimator (QMLE) of an ACD model (Engle and Russell, 1998). Also, as noted earlier, some tests for the standardized innovation distribution assume correct specification for conditional expected duration. Our tests have several appealing features. First, it can detect neglected linear and nonlinear dynamic structure in conditional expected duration. Nonlinear features are not uncommon in high-frequency financial data. Since market activities are often driven by the arrival of news, it is possible that the trading dynamics measured by intraday transaction durations are different between heavy and thin trading periods. Engle and Russell (1998) are perhaps the first to recognize the need to account for nonlinearity in modelling durations of financial events. They use a simple test to detect nonlinearity and find that conditional durations of a trade are overpredicted by a linear ACD model after shortest or longest durations. This suggests that the standard linear ACD model of Engle and Russell (1998) cannot fully capture nonlinear dynamics in durations. Zhang *et al.* (2001) also document that the dynamics of a short duration regime, which is associated with informed trading, is different from the dynamics of a long duration regime, which is associated with uninformed trading. They find that the short duration regime is characterized by wider spreads, larger volume and higher volatility, all of which proxy for informed trading. There have been various nonlinear extensions of Engle and Russell's (1998) linear ACD model. These include fractionally integrated ACD models of Jasiak (1999), log-ACD models of Bauwen and Giot (2000), Box–Cox ACD models of Dufour and Engle (2000) and Hautsch (2003), threshold ACD (TACD) models of Zhang *et al.* (2001), Markov-switching ACD models of Hujer *et al.* (2002), smooth transition ACD models of Meitz and Teräsvirta (2006) and asymmetric ACD models of Fernandes and Grammig (2006). Each nonlinear ACD model can capture some nonlinear duration features. However, given these increasing nonlinear specifications, it is unclear which type of ACD model would fit a financial duration data adequately. Therefore, it is important to have a generally applicable test for ACD models that can detect a wide range of neglected linear and nonlinear dynamics in durations.

Most existing works in the ACD literature assume that standardized innovations are i.i.d. Such an assumption is convenient but may not be suitable for nonlinear ACD models. For example, a regime-switching ACD model assumes that depending on the state of the latent information regime corresponding to heavier or thinner trading periods, trade durations follow different data-generating mechanisms (i.e. fast and slow regimes have different dynamics). Thus, it is more appropriate to assume that standardized innovations follow a mixture distribution with unit mean but time-varying higher-order conditional moments (e.g. Hujer, *et al.*, 2002, appendix A.2). In fact, one implication of the i.i.d. innovations assumption is that the ACD model does not allow for independent variation of the conditional mean and dispersion as higher-order conditional moments are solely linked to the conditional mean. Ghysels *et al.* (2004) argue that this is a very restrictive assumption, especially in analysis of market liquidity. Drost and Werker (2004) show that for commonly used ACD models, the assumption of i.i.d. innovations is too restrictive and inappropriate to describe financial durations accurately (see also Pacurar, 2006). They relax the i.i.d. innovations assumption and consider its impact on semi-parametric estimation efficiency of an ACD model. Zhang *et al.* (2001) also relax the i.i.d. innovations assumption via a regime-switching model. It is important to take into account the impact of serial dependence in the higher-order conditional moments of standardized innovations when constructing a test for ACD models. Failure to do so may cause incorrect Type I errors, as illustrated in our simulation study.

Our tests are robust to time-varying conditional dispersion and higher-order conditional moments of unknown form. We use a generalized spectral derivative approach. The generalized spectrum, originally proposed in Hong (1999), is a frequency domain tool for nonlinear time-series analysis. It is essentially a spectral analysis of time series transformed via the characteristic function.

The generalized spectrum itself is not suitable for testing conditional expected duration models, because it can capture serial dependence not only in the conditional mean but also in higher-order conditional moments. However, using a suitable partial derivative of the generalized spectrum, we can construct tests that solely focus on the conditional expected duration dynamics. Thanks to the use of the characteristic function, our tests can detect a wide class of neglected linear and nonlinear dynamic structures in conditional expected duration. Also, thanks to the use of the spectral analysis, the proposed tests can check a large number of lags without suffering from the curse of dimensionality. This is particularly appealing in the present context because most ACD models are non-Markovian, where the conditioning observable information set is infinite-dimensional containing an infinite number of lags (i.e. the entire past history).

Our tests only require estimation of the null ACD model and have a convenient null asymptotic N(0,1) distribution. Unlike the Box–Pierce–Ljung portmanteau test, parameter estimation uncertainty has no impact on the asymptotic distribution of the proposed tests for ACD models. However, the impact of parameter estimation uncertainty is not trivial in finite samples, as revealed in the following simulation study. To alleviate it, we adopt Wooldridge's (1990a) device to our context, which can effectively remove the impact of parameter estimation uncertainty. By running an increasing sequence of auxiliary regressions, we can reduce the impact of parameter estimation uncertainty for the generalized spectral derivative tests of ACD models. As a result, the finite sample distribution of the tests becomes robust to parameter estimation uncertainty to some extent. Arguably, the reasonable size performance with the convenient asymptotic N(0,1) distribution is one of the most appealing properties of our tests from a practitioner's point of view. In particular, the bootstrap procedure can be avoided, which would be rather computationally expensive for massive high-frequency financial time-series data especially when the standardized innovations have time-varying conditional higher-order moments of unknown form.

We note that Hong and Lee (2007) also use Wooldridge's (1990a) device to remove the impact of parameter estimation uncertainty in testing a time-series regression model with additive regression errors. Because we explicitly explore the multiplicative error structure in an ACD model and can eliminate conditional dispersion clustering completely when the standardized innovations are i.i.d., our approach here is expected to give better size and power for the proposed tests (see Section 6 for more discussion). The multiplicative error structure of ACD also results in a different form of regressors in running Wooldridge's (1990a) auxiliary regressions. Moreover, the moment condition on the multiplicative errors is much weaker.

Although our tests are motivated by checking the adequacy of ACD models, they are readily applicable to strictly stationary time-series models with a multiplicative error structure with a non-negative conditional expectation. Such models are often used to characterize the dynamics of non-negative time-series processes. Non-negative time-series are common in finance. Examples include the volume of shares traded over a period, the ask-bid price spread and the number of trades in a period. For instance, our tests can be used to check the conditional autoregressive range (CARR) model proposed by Chou (2005) for the high–low price spread of stock prices. As such, the main contribution of this article is to propose a general framework to test multiplicative error models for non-negative processes. The methodology is related to Hong and Lee (2005, 2007), which proposed a framework for testing the martingale difference sequence (m.d.s.) hypothesis in a sense that the moment restriction we are testing is a m.d.s. property of standardized errors implied by correct specification of a conditional mean model. However, Hong and Lee (2005) consider additive errors whereas we used standardized multiplicative errors here. As explained next, when testing multiplicative error models including ACD models, the tests based on additive model residuals may be asymptotically less powerful because of the existence of conditional heteroskedasticity of unknown form in the additive model residuals even when the standardized errors are i.i.d. Our use of standardized multiplicative model residuals avoid such an undesired feature.

Section 2 introduces hypotheses of interest and the testing approach. We propose the generalized spectral derivative tests in Section 3, and derive their asymptotic normal distribution in Section 4. Section 5 considers a finite sample correction to remove parameter estimation uncertainty. Section 6 investigates the asymptotic power property of the tests. Section 7 examines their finite sample performance via Monte Carlo experiments. Section 8 concludes. All mathematical proofs are collected in the appendix. Throughout, we denote C for a generic bounded constant, A^* for the complex conjugate of A, ReA for the real part of A and $||A||$ for the Euclidean norm of A. All limits are taken as the sample size $n \to \infty$. The GAUSS code to implement our tests is available from the authors on request.

2. HYPOTHESES OF INTEREST AND APPROACH

2.1. Hypotheses of interest

In subsequent sections, we develop a general theory to test model specification of a non-negative time-series process $\{Y_i\}_{i=1}^{\infty}$ of the multiplicative form

$$Y_i = \psi_i^0 \varepsilon_i, \tag{1}$$

where ε_i is a non-negative innovation, $\psi_i^0 \equiv E(Y_i \mid I_{i-1})$ is the conditional expectation, I_{i-1}, is the information set that contains lagged values of Y_i and possibly other lagged observable variables available at time indexed by $i-1$, and this information set is increasing in i. Non-negative time series processes are common in time-series analysis and occur in many applied areas, such as economics and finance. A prime example is a point process $\{t_i, i = 1, 2, \ldots\}$, a sequence of strictly increasing random variables, corresponding to arrival times of events of interest, such as transactions. Here, one may be interested in the dynamics of $Y_i = t_i - t_{i-1}$, the elapsed time between two consecutive events occurring at times t_i and t_{i-1} respectively. Other examples of Y_i include the volume of shares over a 10-minute period, the high price minus the low price over a time period, the ask price minus the bid price and the number of trades in a period (Engle, 2002).

The multiplicative innovation ε_i is also called a standardized innovation because $\varepsilon_i = Y_i/\psi_i^0$. By construction, $E(\varepsilon_i \mid I_{i-1}) = 1$ almost surely. It is often assumed in the literature that $\{\varepsilon_i\}$ is i.i.d. For example, $\{\varepsilon_i\}$ can follow an i.i.d. sequence of standard exponential or Weibull random variables, as in the ACD model of Engle and Russell (1998). In this case, all past information enters the current duration Y_i via the conditional expected duration ψ_i^0, which captures the full dynamics of Y_i. However, the i.i.d. assumption for $\{\varepsilon_i\}$ may be too restrictive in practice. It rules out the possibility that the conditional dispersion of $\{\varepsilon_i\}$ is time-varying. As Engle (2000) points out, $\{\varepsilon_i\}$ may follow a non-negative distribution with a unit mean and time-varying variance, and there are many such candidates. One example is

$$\begin{cases} \varepsilon_i = \exp(\sqrt{h_i} z_i)/\exp(\frac{1}{2}h_i), \\ z_i \sim \text{i.i.d. } N(0,1), \end{cases} \tag{2}$$

where $h_i = h(I_{i-1})$, $E(\varepsilon_i \mid I_{i-1}) = 1$ and $\text{var}(\varepsilon_i \mid I_{i-1}) = \exp(h_i) - 1$. Liu *et al.* (2006) document that the standardized innovation $\{\varepsilon_i\}$ is not i.i.d. for both intraday Eurodollars and Japanese Yens. In fact, many empirical applications allow $\{\varepsilon_i\}$ having different dispersions across different regimes. In such scenarios, $\{\varepsilon_i - 1\}$ is an m.d.s. but not i.i.d.

The flexibility of modelling the conditional expectation of $\{Y_i\}$ arises from various choices of models for the conditional expectation, ψ_i^0, and the probability density function of ε_i. Conditional expectation ψ_i^0 contains useful information about the dynamics of $\{Y_i\}$. For example, in ACD modelling, long durations indicate lack of trading activities, which signifies a period of no new information. To capture the dynamics of conditional expectation, practitioners often use a parametric model for ψ_i^0. An example is Engle and Russell's (1998) linear ACD(p,q) model. Suppose $\psi(I_{i-1},\theta)$, $\theta \in \Theta \subset \mathbb{R}^r$, is a parametric model for ψ_i^0, where Θ is a finite-dimensional parameter space. We say that $\psi(I_{i-1}, \theta)$ is correctly specified for ψ_i^0 if

$$\mathbb{H}_0 : \psi(I_{i-1}, \theta_0) = \psi_i^0 \text{ almost surely for some } \theta_0 \in \Theta \subset \mathbb{R}^p.$$

Alternatively, we say that $\psi(I_{i-1},\theta)$ is misspecified for ψ_i^0 if

$$\mathbb{H}_A : \text{ There exists no } \theta \in \Theta \text{ such that } \psi(I_{i-1}, \theta) = \psi_i^0 \text{ almost surely.}$$

Our goal is to develop tests for $\mathbb{H}_0$ vs. $\mathbb{H}_A$ that can detect a wide range of misspecifications in $\psi(I_{i-1}, \theta)$ while being robust to time-varying higher-order conditional moments of ε_i.

2.2. Generalized spectral derivative analysis

In practice, $\{Y_i\}$ is often a non-Markovian process, as is the case for almost all ACD models considered in the literature. As a result, the conditioning information set I_{i-1} is infinite-dimensional (i.e. dating back to the infinite past) or its dimension grows with time t_i. This poses a challenge in testing the model $\psi(I_{i-1}, \theta)$, because of the curse of dimensionality. To avoid it, we will propose a non-parametric test of $\mathbb{H}_0$ using a suitable partial derivative of Hong's (1999) generalized spectrum. Define the standardized model error

$$\varepsilon_i(\theta) \equiv \frac{Y_i}{\psi(I_{i-1}, \theta)}, \quad \theta \in \Theta \subset \mathbb{R}^p. \tag{3}$$

Then $\mathbb{H}_0$ holds if and only if $E[\varepsilon_i(\theta_0) \mid I_{i-1}] = 1$ a.s. for some $\theta_0 \in \Theta$. This implies $E[\varepsilon_i(\theta_0) \,|\, I_{i-1}^{\varepsilon}] = 1$ a.s., where $I_{i-1}^{\varepsilon} \equiv \{\varepsilon_{i-1}(\theta_0), \varepsilon_{i-2}(\theta_0), \ldots\}$. It forms a basis for testing $\mathbb{H}_0$. We note that one could also test $\mathbb{H}_0$ by using the additive error $\xi_i = Y_i - \psi(I_{i-1},\theta)$ (see Hong and Lee, 2005, 2007) rather than the multiplicative error ε_i in (3). However, the test based on $\{\xi_i\}$ may result in an asymptotically less powerful test than a test based on $\{\varepsilon_i\}$ because $\{\xi_i = \varepsilon_i[\psi_i^0 - \psi(I_{i-1}, \theta)]\}$ is conditionally heteroskedastic even when $\{\varepsilon_i\}$ is i.i.d. (see Section 6 for more discussion). In addition, the use of $\{\varepsilon_i\}$ rather than $\{\xi_i\}$ allows weaker moment conditions on the data-generating process (DGP). In particular, we allow integrated ACD models which are strictly but not weakly stationary [an integrated ACD(1,1) model is given as follows: $\psi_i = \alpha + \beta\psi_{i-1} + \gamma Y_{i-1}$, with $\beta + \gamma = 1$]. In this case, $\{\varepsilon_i\}$ is still weakly stationary but $\{\xi_i\}$ is not.

For notational economy, we put $\varepsilon_i \equiv \varepsilon_i(\theta_0)$, where θ_0 is the probability limit of some parameter estimator $\hat{\theta}$ and θ_0 satisfies the condition that $\psi_i^0 = \psi(I_{i-1}, \theta_0)$ almost surely under $\mathbb{H}_0$. Li and Yu (2003) propose a portmanteau diagnostic test for $\mathbb{H}_0$ using a modified Box–Pierce (1970) type test statistic based on finitely many sample autocorrelations of $\{\varepsilon_i\}$, in a spirit similar to Li and Mak (1994). The modification takes into account the impact of parameter estimation uncertainty in the ACD model. The resulting test statistic has a convenient asymptotically valid chi-squared distribution under $\mathbb{H}_0$, and has power against dynamic misspecification (i.e. misspecification in lag order structure). The test also has good power against many nonlinear ACD alternatives, although it may miss some important nonlinear ones because of the use of the autocovariance function of $\{\varepsilon_i\}$. Meitz and Teräsvirta (2006) also propose a class of LM-type tests against some specific ACD alternatives. These tests are most powerful against the assumed alternatives.

We note that there are tests for conditional moment restrictions with respect to an infinite information set in the literature. For example, Escanciano and Velasco (2006) proposed a test for the martingale hypothesis for raw data, and Escanciano (2006) proposed a specification test for parametric conditional mean models. The former is not directly applicable for testing ACD models since they do not consider estimated model residuals, whereas the sampling variation of parameter estimation affects the asymptotic distribution of their test statistic. The latter can be viewed as an alternative to Hong and Lee (2005) and could be used to test ACD models. Like Hong and Lee (2005), however, Escanciano's (2006) test is based on additive model residuals; thus, it may be less powerful than the test based on standardized errors as the test proposed in the present article. Also, the non-standard limiting distribution of his test requires computationally costly resampling.

Because no prior information about the true alternative is usually available to practitioners, it is highly desirable to develop complementary tests for ACD models that do not require the knowledge of the alternative and have reasonable power against a wide range of neglected linear and nonlinear ACD alternatives. We now develop a class of such tests using a generalized spectral derivative approach. Suppose $\{\varepsilon_i\}$ is a strictly stationary process with marginal characteristic function $\varphi(u) \equiv E(e^{\mathbf{i}u\varepsilon_i})$ and pairwise joint characteristic function $\varphi_j(u,v) \equiv E(e^{\mathbf{i}u\varepsilon_i+\mathbf{i}v\varepsilon_{i-|j|}})$, where $\mathbf{i} \equiv \sqrt{-1}$, $u, v \in \mathbb{R}$, and $j = 0, \pm 1, \ldots$. The basic idea of the generalized spectrum in Hong (1999), tailored to the present context, is to consider the spectrum of the transformed series $\{e^{\mathbf{i}u\varepsilon_i}\}$, which is defined as

$$f(\omega,u,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(u,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi], \ \ u,v \in \mathbb{R}, \tag{4}$$

where ω is the frequency, and $\sigma_j(u,v) \equiv \mathrm{cov}(e^{\mathbf{i}u\varepsilon_i}, e^{\mathbf{i}v\varepsilon_{i-|j|}})$ is the covariance function of the transformed series. The function $f(\omega, u, v)$ is well defined when

$$\sum_{j=-\infty}^{\infty}\sup_{(u,v)\in\mathbb{R}^2}|\sigma_j(u,v)| < \infty,$$

which holds if $\{\varepsilon_i\}$ is an α-mixing process with α-mixing coefficients satisfying the restriction that

$$\sum_{j=-\infty}^{\infty}\alpha(j)^{(v-1)/v} < \infty$$

for some $v > 1$ (see Hong, 1999). It can capture any type of pairwise serial dependence in $\{\varepsilon_i\}$, that is dependence between ε_i and ε_{i-j} for any lag $j \neq 0$, including nonlinear serial dependence with zero autocorrelation. This is analogous to the higher-order spectra (Brillinger, 1965; Brillinger and Rosenblatt, 1967a, 1967b). Unlike the higher-order spectra, however, $f(\omega, u, v)$ does not require the existence of any moment of $\{\varepsilon_i\}$. When $E(\varepsilon_i^2)$ exists, we can obtain the conventional power spectrum from a partial derivative of $f(\omega,u,v)$ at $(u,v) = (0,0)$:

$$-\frac{\partial^2}{\partial u\partial v}f(\omega,u,v)\,|_{(u,v)=(0,0)} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\mathrm{cov}(\varepsilon_i,\varepsilon_{i-|j|})e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi],$$

where interchanging differentials and expectation is valid given

$$\sum_{j=-\infty}^{\infty}|\mathrm{cov}(\varepsilon_i,\varepsilon_{i-|j|})| < \infty.$$

For this reason, $f(\omega, u, v)$ is called the generalized spectrum of $\{\varepsilon_i\}$.

As is well known, the interpretation of spectral analysis is more difficult for nonlinear time series than for linear time series. Unlike the power spectrum, the higher-order spectra have no physical interpretation (i.e. energy decomposition over frequencies). This is also true of $f(\omega, u, v)$. However, the basic idea of characterizing cyclical dynamics still applies: $f(\omega, u, v)$ is useful when searching for linear or nonlinear cyclical movements. A strong cyclicity of data can be linked with a strong serial dependence in $\{\varepsilon_i\}$ that may not be captured by the autocorrelation function of $\{\varepsilon_i\}$. The generalized spectrum $f(\omega, u, v)$ can capture such nonlinear cyclical patterns by displaying distinct spectral peaks. This can be seen from the Taylor series expansion of $f(\omega,\cdot,\cdot)$ around the origin (0,0):

$$f(\omega,u,v) = \sum_{m=0}^{\infty}\sum_{l=0}^{\infty}\frac{(\mathbf{i}u)^m(\mathbf{i}v)^l}{m!l!}\left[\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\mathrm{cov}(\varepsilon_i^m,\varepsilon_{i-|j|}^l)e^{-\mathbf{i}j\omega}\right], \quad \omega \in [-\pi,\pi], \ \ u,v \in \mathbb{R},$$

which holds under suitable regularity conditions. Now suppose $\{\varepsilon_i\}$ is a white noise [$\mathrm{cov}(\varepsilon_i, \varepsilon_{i-j}) = 0$ for all $j \neq 0$] but has a stochastic cyclical pattern in dispersion clustering. Then the power spectrum will miss such dispersion clustering, but $f(\omega, u, v)$ can effectively capture it. More generally, $f(\omega, u, v)$ can capture cyclical dynamics in the conditional distribution of $\{\varepsilon_i\}$, including those in the tail clustering of the distribution.

Correct specification for ψ_i^0 is equivalent to the condition that $E(\varepsilon_i \mid I_{i-1}) = 1$ a.s. It is possible that $\{\varepsilon_i\}$ is not i.i.d. under $\mathbb{H}_0$, as is illustrated in (2). The generalized spectrum $f(\omega, u, v)$ itself is not suitable for testing $\mathbb{H}_0$, because it can capture serial dependence in not only conditional mean but also higher-order conditional moments. In other words, it may incorrectly reject $\mathbb{H}_0$ because of the existence of serial dependence in higher-order conditional moments of ε_i rather than the violation of $\mathbb{H}_0$.

However, just as the characteristic function can be differentiated to generate various moments of $\{\varepsilon_i\}$, $f(\omega, u, v)$ can be differentiated to capture serial dependence in various moments of $\{\varepsilon_i\}$. To focus on and only on the departures from $E(\varepsilon_i \mid I_{i-1}) = 1$, one can use the partial derivative

$$f^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j^{(1,0)}(0,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi], \quad v \in \mathbb{R}, \tag{5}$$

where

$$\sigma_j^{(1,0)}(0,v) \equiv \frac{\partial}{\partial u}\sigma_j(u,v)\,|_{u=0} \; = \; \mathrm{cov}(\mathbf{i}\varepsilon_i, e^{\mathbf{i}v\varepsilon_{i-|j|}}).$$

Note that $\sigma_j^{(1,0)}(0,v) = 0$ for all $v \in \mathbb{R}$ if and only if $E(\varepsilon_i \mid \varepsilon_{i-|j|}) = 1$ given the boundedness of the complex-valued exponential function $e^{\mathbf{i}v\varepsilon_{i-|j|}}$ and $E|\varepsilon_t| < \infty$ (see Stinchcombe and White, 1998, Thm. 2.3). The function $E(\varepsilon_i \mid \varepsilon_{i-|j|})$ is called the autoregression function of $\{\varepsilon_i\}$ in nonlinear time series analysis (Tong, 1990, p. 8) and can capture linear and nonlinear dependences in the conditional mean of $\{\varepsilon_i\}$, including the processes with zero autocorrelation. Therefore, $\sigma_j^{(1,0)}(0,v)$, or equivalently $f^{(0,1,0)}(\omega, 0, v)$, ideally suits for testing $\psi(I_{i-1}, \theta)$. Moreover, although $E(\varepsilon_i \mid \varepsilon_{i-|j|})$ and $\sigma_j^{(1,0)}(0,v)$ are equivalent measures, the use of $\sigma_j^{(1,0)}(0,v)$ rather than $E(\varepsilon_i \mid \varepsilon_{i-|j|})$ avoids smoothed non-parametric estimation.

Under $\mathbb{H}_0$, the generalized spectral derivative $f^{(0,1,0)}(\omega,0,v)$ becomes a 'flat spectrum':

$$f_0^{(0,1,0)}(\omega,0,v) = \frac{1}{2\pi}\sigma_0^{(1,0)}(0,v) \quad \text{for all } \omega \in [-\pi,\pi], \quad \text{and } v \in \mathbb{R}.$$

Thus, one can test $\mathbb{H}_0$ vs. $\mathbb{H}_A$ by comparing two consistent estimators for $f^{(0,1,0)}(\omega,0,v)$ and $f_0^{(0,1,0)}(\omega,0,v)$ respectively. Under $\mathbb{H}_0$, these estimators converge to the same limit. If they converge to different limits, there exists evidence against $\mathbb{H}_0$. Note that we always have a flat spectrum under $\mathbb{H}_0$ even if there exists conditional dispersion clustering [i.e. $\mathrm{cov}(\varepsilon_i^2, \varepsilon_{i-j}^2) \neq 0$ for some $j \neq 0$]. This provides a basis for constructing tests for $\mathbb{H}_0$ that are robust to time-varying higher-order conditional moments in $\{\varepsilon_i\}$.

3. GENERALIZED SPECTRAL DERIVATIVE TESTS

Because $\{\varepsilon_i\}$ is not observed, we need to use an estimated standardized model residual

$$\hat{\varepsilon}_i \equiv \frac{Y_i}{\psi(I_{i-1}^{\dagger}, \hat{\theta})}, \quad i = 1, \ldots, n, \tag{6}$$

where $I_{i-1}^{\dagger}$ is the feasible information set observed at time t_i−1 that may involve some assumed initial values. For example, consider an ACD(1,1) model $Y_i = \psi_i\varepsilon_i$, where $\psi_i = \alpha + \beta\psi_{i-1} + \gamma Y_{i-1}$. Here, the infeasible information set $I_{i-1} = \{Y_{i-1}, Y_{i-2}, \ldots, Y_1, Y_0, \ldots\}$ contains the entire past history $\{Y_s, s < i\}$ dating back to the infinite past. On the other hand, $I_{i-1}^{\dagger} = \{Y_{i-1}, Y_{i-2}, \ldots, Y_1, \bar{Y}_0, \bar{\psi}_0\}$, where $\bar{Y}_0, \bar{\psi}_0$ are some assumed initial values for Y_0, ψ_0 respectively.

In (6), any $\sqrt{n}$-consistent parameter estimator $\hat{\theta}$ based on a random sample $\{Y_i\}_{i=1}^n$ of size n can be used. An example of $\hat{\theta}$ is the QMLE of Engle and Russell (1998) in estimating an ACD model, which is based on the assumption that $\{\varepsilon_i\} \sim$ i.i.d. exp(1):

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n}\left[\ln\psi(I_{i-1}^{\dagger}, \theta) + \frac{Y_i}{\psi(I_{i-1}^{\dagger}, \theta)}\right]. \tag{7}$$

Engle and Russell (1998) show that $\hat{\theta}$ in (7) is consistent for θ_0 under $\mathbb{H}_0$ even if $\{\varepsilon_i\}$ is not i.i.d. exp(1), although it is not asymptotically most efficient. More generally, Drost and Werker (2004) show that QMLE is consistent when it is based on the standard Gamma family.

With $\{\hat{\varepsilon}_i\}_{i=1}^n$, one can estimate $f^{(0,1,0)}(\omega, 0, v)$ by a non-parametric smoothed kernel estimator

$$\hat{f}^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=1-n}^{n-1}\left(\frac{1-|j|}{n}\right)^{1/2} k\left(\frac{j}{p}\right)\hat{\sigma}_j^{(1,0)}(0,v)e^{-\mathbf{i}j\omega}, \quad \omega \in [-\pi,\pi], \;\; v \in \mathbb{R},$$

where

$$\hat{\sigma}_j^{(1,0)}(0,v) = \frac{1}{n-|j|}\sum_{i=|j|+1}^{n} \mathbf{i}(\hat{\varepsilon}_i - 1)\hat{\phi}_{i-|j|}(v), \tag{8}$$

$\hat{\phi}_{i-|j|}(v) = e^{\mathbf{i}v\hat{\varepsilon}_{i-|j|}} - \hat{\varphi}(v)$ and $\hat{\varphi}(v) = n^{-1}\sum_{i=1}^{n} e^{\mathbf{i}v\hat{\varepsilon}_i}$. One could replace unity in (8) by the sample mean of $\{\hat{\varepsilon}_i\}$. Here, $p \equiv p(n)$ is a bandwidth that grows with the sample size n, and $k : \mathbb{R} \to [-1,1]$ is a symmetric kernel that assigns weights to various lags. Examples of $k(\cdot)$ include the Bartlett, Daniell, Parzen and quadratic spectral kernels (e.g. Priestley, 1981, p. 442). The factor $(1-|j|/n)^{1/2}$ is a finite-sample correction. It could be replaced by unity.

To estimate the flat spectral derivative $f_0^{(0,1,0)}(\omega,0,v)$, we use the estimator

$$\hat{f}_0^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\hat{\sigma}_0^{(1,0)}(0,v), \quad \omega \in [-\pi,\pi], \;\; v \in \mathbb{R}.$$

Under $\mathbb{H}_0$, $\hat{f}^{(0,1,0)}(\omega,0,v)$ and $\hat{f}_0^{(0,1,0)}(\omega,0,v)$ converge to the same limit. Under $\mathbb{H}_A$, they generally converge to different limits. Thus, we can test $\mathbb{H}_0$ based on the comparison of $\hat{f}^{(0,1,0)}(\omega,0,v)$ and $\hat{f}_0^{(0,1,0)}(\omega,0,v)$ via a divergence measure (e.g. L_2-norm). Any significant difference between them will be evidence against $\mathbb{H}_0$.

3.1. Tests under martingale difference sequence standardized innovations

An important feature of $\mathbb{H}_0$ is that it is silent about the higher-order conditional moments of $\{\varepsilon_i\}$. In a Markov-chain regime-switching ACD model, for example, the innovation $\{\varepsilon_i\}$ may have different dispersions and time-varying higher-order conditional moments across different regimes. Thus, $\{\varepsilon_i - 1\}$ is m.d.s. but not i.i.d. although it may be i.i.d. within each regime. Hence, the assumption of i.i.d. innovations is too restrictive because it ignores the impact of dispersion clustering in $\{\varepsilon_i\}$ across different regimes. This may result in an incorrect Type I error. Therefore, it is highly desirable to develop tests for $\mathbb{H}_0$ that are robust to time-varying higher-order conditional moments in $\{\varepsilon_i\}$.

A class of tests with such an appealing robust property can be constructed by comparing $\hat{f}^{(0,1,0)}(\omega,0,v)$ and $\hat{f}_0^{(0,1,0)}(\omega,0,v)$ via the quadratic form

$$\begin{aligned} &n\int\int_{-\pi}^{\pi}|\hat{f}^{(0,1,0)}(\omega,0,v)-\hat{f}_0^{(0,1,0)}(\omega,0,v)|^2\mathrm{d}\omega\mathrm{d}W(v)\\ &=\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)(n-j)\int\left|\hat{\sigma}_j^{(1,0)}(0,v)\right|^2\mathrm{d}W(v), \end{aligned} \tag{9}$$

where the equality follows by Parseval's identity. The resulting test statistic is a properly standardized version of (9):

$$\hat{M}_1(p)\equiv\frac{\left[\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)(n-j)\int\left|\hat{\sigma}_j^{(1,0)}(0,v)\right|^2\mathrm{d}W(v)-\hat{C}_1(p)\right]}{\sqrt{\hat{D}_1(p)}}, \tag{10}$$

where $W:\mathbb{R}\to\mathbb{R}^+$ is a non-decreasing function that weighs sets of v symmetric around 0 equally, and the centering and scaling factors are, respectively,

$$\hat{C}_1(p)=\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)\frac{1}{n-j}\sum_{i=j+1}^{n-1}(\hat{\varepsilon}_i^2-1)\int\left|\hat{\phi}_{i-j}(v)\right|^2\mathrm{d}W(v),$$

$$\hat{D}_1(p)=2\sum_{j=1}^{n-2}\sum_{l=1}^{n-2}k^2\left(\frac{j}{p}\right)k^2\left(\frac{l}{p}\right)\int\int\left|\frac{1}{n-\max(j,l)}\sum_{i=\max(j,l)+1}^{n}(\hat{\varepsilon}_i^2-1)\hat{\phi}_{i-j}(v)\hat{\phi}_{i-l}(v')\right|^2\mathrm{d}W(v)\mathrm{d}W(v').$$

Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. An example of $W(\cdot)$ is the N(0,1) CDF, which is commonly used in the empirical characteristic function literature. The factors $\hat{C}_1(p)$ and $\hat{D}_1(p)$ are the approximate mean and variance of the quadratic form in (9). In deriving the forms of $\hat{C}_1(p)$ and $\hat{D}_1(p)$, we have exploited the implication of $\mathbb{H}_0: E(\varepsilon_i|I_{i-1})=1$, and we have taken into account the impact of conditional dispersion clustering and time-varying higher-order conditional moments in $\{\varepsilon_i\}$. As a result, $\hat{M}_1(p)$ is robust to dispersion clustering and time-varying higher-order conditional moments of unknown form, as can occur in a threshold or regime-switching ACD model. In fact, we conjecture that $\hat{M}_1(p)$ is still applicable even if $\{\varepsilon_i\}$ displays unconditional heteroskedasticity [i.e. var(ε_i) differs from i]. Note that both $\hat{C}_1(p)$ and $\hat{D}_1(p)$ grow to infinity at a rate of p as $p\to\infty$, $p/n\to 0$ (see the Appendix for details).

3.2. Tests under i.i.d. standardized innovations

Although the robust test $\hat{M}_1(p)$ is applicable no matter whether $\{\varepsilon_i - 1\}$ is i.i.d. or m.d.s., we can obtain a simpler test statistic with better finite sample performance when $\{\varepsilon_i\}$ is i.i.d. In this case, we can define a simpler test statistic

$$\hat{M}_0(p)\equiv\frac{\left[\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)(n-j)\int\left|\hat{\sigma}_j^{(1,0)}(0,v)\right|^2\mathrm{d}W(v)-\hat{C}_0(p)\right]}{\sqrt{\hat{D}_0(p)}}, \tag{11}$$

where the centering and scaling factors now are simplified as follows:

$$\hat{C}_0(p)=(\hat{s}^2-1)\int[1-|\hat{\varphi}(v)|^2]\mathrm{d}W(v)\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right),$$

$$\hat{D}_0(p)=2(\hat{s}^2-1)^2\int\int|\hat{\varphi}(v+v')-\hat{\varphi}(v)\hat{\varphi}(v')|^2\mathrm{d}W(v)\mathrm{d}W(v')\sum_{j=1}^{n-2}k^4\left(\frac{j}{p}\right),$$

with $\hat{s}^2=n^{-1}\sum_{i=1}^n\hat{\varepsilon}_i^2$. In deriving the forms of $\hat{C}_0(p)$ and $\hat{D}_0(p)$, we have exploited the implication of the i.i.d. assumption on $\{\varepsilon_i\}$. As a result, $\hat{C}_0(p)$ and $\hat{D}_0(p)$ are simpler than $\hat{C}_1(p)$ and $\hat{D}_1(p)$ under the m.d.s. case. We emphasize that $\hat{M}_0(p)$ is not a test for the i.i.d. hypothesis of $\{\varepsilon_i\}$. Instead, it is a test for $\mathbb{H}_0$ (conditional expected duration specification) with the auxiliary assumption that $\{\varepsilon_i\}$ is i.i.d. (i.e. the higher-order conditional moments of $\{\varepsilon_i\}$ are constant). Note that if the additive error, $\xi_i = Y_i - \psi(I_{i-1}, \theta)$, was used, such a simple test statistic as $\hat{M}_0(p)$ cannot be obtained, because of the presence of conditional heteroskedasticity in $\{\xi_i\}$ even when $\{\varepsilon_i\}$ is i.i.d.

As shown in Section 4, $\hat{M}_1(p) \to^d N(0,1)$ under $\mathbb{H}_0$ and $\hat{M}_0(p) \to^d N(0,1)$ when $\{\varepsilon_i\}$ is i.i.d. To gain the intuition, we consider an example of $\hat{M}_0(p)$ where the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$ is used, where $\mathbf{1}(\cdot)$ is the indicator function. In this case, the sum in (9) becomes

$$\sum_{j=1}^{p}(n-j)\int\left|\hat{\sigma}_j^{(1,0)}(0,v)\right|^2 \mathrm{d}W(v),$$

and both $\hat{C}_0(p)$ and $\hat{D}_0(p)$ are proportional to p and $2p$ respectively. Because the sequence

$$\left\{(n-j)\int\left|\hat{\sigma}_j^{(1,0)}(0,v)\right|^2 \mathrm{d}W(v)\right\}_{j=1}^{p}$$

is approximately i.i.d. when $\mathbb{H}_0$ holds, we have, by the central limit theorem, that the sum in (9) converges to N(0, 1) after proper centering and scaling. Of course, our formal proof does not rely on this simplistic heuristics.

4. ASYMPTOTIC NULL DISTRIBUTION

To derive the asymptotic distribution of the proposed tests under $\mathbb{H}_0$, we first provide some regularity conditions.

Assumption A.1. *$\{Y_i\}$ is a strictly stationary non-negative time-series process such that $\psi_i^0 \equiv E(Y_i \mid I_{i-1})$ exists a.s., where I_{i-1} is an information set at time t_{i-1} that may contain lagged dependent variables $\{Y_{i-j}, j>0\}$ as well as current and lagged exogenous variables $\{Z_{i-j}, j \geq 0\}$.*

Assumption A.2. *$\psi(I_{i-1},\theta)$ is a parametric model for ψ_i^0, where $\theta \in \Theta \subset \mathbb{R}^r$ is a finite-dimensional parameter and Θ is a parameter space, such that (a) for each $\theta \in \Theta$, $\psi(\cdot,\theta)$ is measurable with respect to I_{i-1}; (b) with probability 1, $\psi(I_{i-1},\cdot)$ is continuously twice differentiable with respect to $\theta \in \Theta$, and for some $\nu > 1$, $E\sup_{\theta\in\Theta}||\frac{\partial}{\partial\theta}\ln\psi(I_{i-1},\theta)||^{4\max(\nu,2)} \leqslant C$, $E\sup_{\theta\in\Theta}||\frac{\partial^2}{\partial\theta\partial\theta'}\ln\psi(I_{i-1},\theta)||^4 \leqslant C$ and $E\sup_{\theta\in\Theta}[\varepsilon_i(\theta)]^{4\max(\nu,4)} \leqslant C$, where $\varepsilon_i(\theta) = Y_i/\psi(I_{i-1},\theta)$.*

Assumption A.3. *Let $I_i^\dagger$ be a feasible observed information set available at time t_i that may contain some assumed initial values. Then*

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left\{E\left[\sup_{\theta\in\Theta}\left|\frac{\psi(I_{i-1}^\dagger,\theta)-\psi(I_{i-1},\theta)}{\psi(I_{i-1}^\dagger,\theta)}\right|\right]^8\right\}^{1/8} \leqslant C.$$

Assumption A.4. *$\hat{\theta}-\theta_0 = O_P(n^{-1/2})$, where $\theta_0 \in$ int (Θ) satisfies the condition that $\psi_i^0 = \psi(I_{i-1},\theta_0)$ almost surely under $\mathbb{H}_0$.*

Assumption A.5. *Put $\varepsilon_i \equiv \varepsilon_i(\theta_0) = Y_i/\psi(I_{i-1},\theta_0)$ and $G_i \equiv \frac{\partial}{\partial\theta}\ln\psi(I_{i-1},\theta_0)$, where θ_0 is as in Assumption A.4. Then $\{\varepsilon_i, G_i'\}'$ is a strictly stationary α-mixing process with α-mixing coefficient $\alpha(j)$ satisfying*

$$\sum_{j=-\infty}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} < \infty$$

for $\nu > 1$ be as in Assumption A.2.

Assumption A.6. *$k : \mathbb{R} \to [-1,1]$ is symmetric around 0, and is continuous at 0 and all points except a finite number of points, with $k(0) = 1$ and $|k(z)| \leqslant C|z|^{-b}$ as $z\to\infty$ for some $b > 3$.*

Assumption A.7. *$W : \mathbb{R} \to \mathbb{R}^+$ is non-decreasing and weighs sets symmetric around 0 equally, with $\int_{-\infty}^{\infty} v^4\,\mathrm{d}W(v) \leq C$.*

Assumption A.8. *For each sufficiently large integer q, there exists a strictly stationary non-negative process $\{\varepsilon_{q,i}\}$ such that as $q \to \infty$, $\varepsilon_{q,i}$ is independent of I_{i-q-1}, $E(\varepsilon_{q,i} \mid I_{i-1}) = 1$ a.s., $E(\varepsilon_i-\varepsilon_{q,i})^4 \leqslant Cq^{-2\kappa}$ for some constant $\kappa \geq 1$.*

Assumption A.1 imposes a strict stationarity condition on $\{Y_i\}$. Assumption A.2 is a set of smoothness and moment conditions on the model $\psi(I_{i-1},\theta)$. It covers many stationary linear and nonlinear ACD models. Assumption A.3 is a condition on the truncation of information set I_{i-1}, which usually contains the information dating back to the very remote past and so may not be completely observable. Because of the truncation, one may have to assume some initial values in estimating the model $\psi(I_{i-1},\theta)$. Assumption A.3 ensures that the use of initial values, if any, has no impact on the limiting distributions of $\hat{M}_1(p)$ and $\hat{M}_0(p)$. For instance, consider an ACD(1,1) model:

$$\psi_i = \alpha + \beta\psi_{i-1} + \gamma Y_{i-1},$$

where $\psi_i = \psi(I_{i-1},\theta), \theta = (\alpha,\beta,\gamma)', \alpha > 0, 0 \leq \beta \leq \bar{\beta} < 1$ and $0 \leq \gamma \leq \bar{\gamma} < 1$. Here, we have $I_{i-1} = \{Y_{i-1}, Y_{i-2},\ldots,Y_1, Y_0,\ldots\}$ but $I^{\dagger}_{i-1} = \{Y_{i-1}, Y_{i-2}, \ldots, Y_1, \bar{Y}_0, \bar{\psi}_0\}$, and $\bar{Y}_0, \bar{\psi}_0$ are initial values assumed for Y_0,ψ_0 respectively. By recursive substitution, we can show

$$\sum_{i=1}^{n}\left\{E\left[\sup_{\theta\in\Theta}\left|\psi(I^{\dagger}_{i-1},\theta) - \psi(I_{i-1},\theta)\right|\right]^8\right\}^{1/8} \leqslant \frac{1}{1-\bar{\beta}}\left\{[E(\bar{\psi}_0^8)]^{1/8} + \frac{\bar{\gamma}}{1-\bar{\beta}}[E(\bar{Y}_0^8)]^{1/8}\right\} \leqslant C.$$

Assumption A.4 requires that $\hat{\theta}$ be a $\sqrt{n}$-consistent estimator under $\mathbb{H}_0$, which need not be asymptotically most efficient, and converges to a constant under the alternative. It can be the QMLE in (7), or an efficient estimator developed in Drost and Werker (2004). We do not need to know the asymptotic expansion structure of $\hat{\theta}$, because the sampling variation in $\hat{\theta}$ does not affect the asymptotic distribution of $\hat{M}_1(p)$. Assumption A.5 imposes a mixing condition on $\{\varepsilon_i, G_i'\}'$, which restricts the degree of temporal dependence in $\{\varepsilon_i, G_i'\}'$. Mixing conditions are convenient for nonlinear time-series analysis. For more discussion on mixing conditions, see (e.g.) White (2001, pp. 46–47).

Assumption A.6 is a regularity condition on the kernel function $k(\cdot)$, which assigns weights to various lags. It includes many commonly used kernels in practice. The condition of $k(0) = 1$ ensures that the asymptotic bias of the smoothed kernel estimator $\hat{f}^{(0,1,0)}(\omega,0,v)$ in (8) vanishes to 0 as $n \to \infty$. The tail condition on $k(\cdot)$ requires that $k(z)$ decay to 0 sufficiently fast as $|z|\to\infty$. It implies that

$$\int_0^\infty (1+z^2)|k(z)|\mathrm{d}z < \infty.$$

This condition rules out the Daniell and quadratic spectral kernels, whose $b = 2$. However, it includes all kernels with bounded support, such as the Bartlett and Parzen kernels, for which $b = \infty$. Assumption A.7 is a condition on the weighting function $W(v)$. It is satisfied by the CDF of any symmetric continuous distribution with a finite fourth moment.

Assumption A.8 is required only under $\mathbb{H}_0$. It assumes that when q is sufficiently large, the innovation ε_i can be approximated by a q-dependent non-negative process $\varepsilon_{q,i}$ arbitrarily well. Horowitz (2003) imposes a similar condition in a different context. Because $E(\varepsilon_i|I_{i-1}) = 1$ under $\mathbb{H}_0$, Assumption A.8 essentially imposes restrictions on the serial dependence in higher-order moments of $\{\varepsilon_i\}$. It holds trivially when $\{\varepsilon_i\}$ is i.i.d. or when $\{\varepsilon_i\}$ is a q_0-dependent process with an arbitrarily large but fixed order q_0. It also covers many non-Markovian innovation processes. This assumption greatly simplifies the proof of the asymptotic normality for the proposed test statistics.

We now derive the asymptotic distributions of the $\hat{M}_1(p)$ and $\hat{M}_0(p)$ tests under $\mathbb{H}_0$.

Theorem 1. *Suppose Assumptions A.1–A.8 hold, and* $p = cn^{\lambda}$ *for* $0 < \lambda < (3 + 1/4b-2)^{-1}$ *and* $0 < c < \infty$. *Then (i)* $\hat{M}_1(p) \to^{d} N(0,1)$ *under* $\mathbb{H}_0$. *(ii) Suppose in addition* $\{\varepsilon_i\}$ *is i.i.d., then* $\hat{M}_0(p) \to^{d} N(0,1)$.

Because a $\sqrt{n}$-consistent estimator $\hat{\theta}$ converges to θ_0 under $\mathbb{H}_0$ faster than the non-parametric estimator $\hat{f}^{(0,1,0)}(\omega,0,v)$ converges to $f^{(0,1,0)}(\omega,0,v)$, the asymptotic distribution of $\hat{M}_1(p)$ is solely determined by the non-parametric estimator $\hat{f}^{(0,1,0)}(\omega,0,v)$. Consequently, unlike the Box–Pierce–Ljung type portmanteau test, parameter estimation uncertainty in $\hat{\theta}$ has no impact on the asymptotic distribution of $\hat{M}_1(p)$, a so-called 'asymptotic nuisance parameter free' property. In other words, the asymptotic distribution of $\hat{M}_1(p)$ remains unchanged when $\hat{\theta}$ is replaced by its probability limit θ_0. This results in a convenient procedure. Only estimated model residuals are needed to compute the test statistics $\hat{M}_1(p)$ and $\hat{M}_0(p)$.

The $\hat{M}_1(p)$ test is applicable no matter whether $\{\varepsilon_i\}$ is i.i.d. under $\mathbb{H}_0$. However, when $\{\varepsilon_i\}$ is i.i.d., we expect that $\hat{M}_0(p)$ has better size than $\hat{M}_1(p)$ in finite samples, because $\hat{M}_0(p)$ exploits the implication of the i.i.d. property of $\{\varepsilon_i\}$ in constructing the centering and scaling factors. Nevertheless, $\hat{M}_0(p)$ is expected to have size distortion when $\{\varepsilon_i\}$ is not i.i.d. under $\mathbb{H}_0$. These are confirmed in our simulation study (see Table 1).

5. REMOVING PARAMETER ESTIMATION UNCERTAINTY

The 'asymptotic nuisance parameter free' property is appealing, but it is not free of cost. Although parameter estimation uncertainty has no impact on the asymptotic distributions of $\hat{M}_1(p)$ and $\hat{M}_0(p)$, it affects their finite sample distributions, particularly when the sample size n is not large. Intuitively, the estimator $\hat{\theta}$ can result in an adjustment of at most a finite number of degrees of freedom to the distributions of $\hat{M}_1(p)$ and $\hat{M}_0(p)$. When the lag order $p\to\infty$ as $n\to\infty$, the impact of $\hat{\theta}$ becomes negligible when normalized by the scaling factor $\hat{D}_1^{1/2}(p)$ or $\hat{D}_0^{1/2}(p)$, which grows to infinity at the rate of $p^{1/2}$. Nevertheless, asymptotic analysis reveals that the asymptotically negligible higher-order terms in $\hat{M}_1(p)$ and $\hat{M}_0(p)$ that are associated with parameter estimation uncertainty vanish to

Table 1. Empirical sizes of tests

	$M_1(\hat{p}_0)$		$M_0(\hat{p}_0)$		$M_1^d(\hat{p}_0)$		$M_0^d(\hat{p}_0)$		$\mathrm{LY}_1(\bar{p})$		$\mathrm{LY}_2(\bar{p})$	
$\bar{p}$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DGP S.1: ACD(1,1) − i.i.d. errors												
$N = 500$												
5	3.7	2.5	4.7	2.9	7.9	4.0	10.0	6.0	7.2	4.2	7.0	3.3
10	3.9	2.7	5.0	3.2	8.6	4.5	10.6	6.9	9.2	4.7	7.4	4.0
20	3.7	2.0	7.2	4.0	6.7	3.4	11.3	7.3	9.0	5.9	8.9	4.8
$N = 1000$												
5	4.2	3.0	4.4	2.9	8.1	4.8	7.5	4.5	6.5	3.2	7.0	3.7
10	4.6	3.0	4.9	2.9	8.6	5.1	7.4	4.3	8.6	5.1	9.1	5.8
20	4.8	2.8	5.9	3.6	8.1	4.8	7.1	2.9	9.9	5.3	8.9	4.9
$N = 2000$												
5	5.0	3.3	5.4	3.4	9.6	5.8	10.8	6.6	7.0	3.8	7.1	3.5
10	5.0	3.4	5.4	3.4	9.8	5.8	10.9	6.6	7.1	3.7	7.8	3.9
20	5.3	2.5	6.8	3.3	9.5	5.9	10.3	7.5	8.7	4.3	8.6	4.6
DGP S.2: ACD(1,1) − non-i.i.d. errors												
$N = 500$												
5	6.0	4.1	17.2	11.6	7.7	5.1	23.4	23.4	29.5	21.3	8.9	5.3
10	6.2	4.4	17.9	12.4	7.8	5.2	24.4	18.0	34.5	29.0	9.1	6.3
20	3.9	3.9	17.7	10.7	5.1	2.9	21.7	15.4	44.0	37.6	11.9	7.7
$N = 1000$												
5	3.7	1.8	17.8	11.7	7.5	3.7	27.8	20.1	41.1	33.9	10.6	5.9
10	3.8	1.8	18.2	12.3	7.9	3.7	28.6	20.3	46.9	39.8	9.6	5.8
20	2.6	1.3	19.1	12.6	5.9	2.6	27.1	18.4	57.7	51.4	12.3	7.3
$N = 2000$												
5	6.0	3.2	25.1	18.4	9.6	6.1	33.2	25.6	56.0	48.5	12.0	8.0
10	6.0	3.2	25.3	18.8	9.7	6.1	33.5	25.8	62.1	54.7	12.8	7.8
20	5.1	3.1	25.0	18.5	8.2	4.8	31.6	25.4	68.0	67.2	13.8	9.4

1000 iterations; $\hat{M}_1(\hat{p}_0), \hat{M}_0(\hat{p}_0)$, generalized spectral tests derived under time-varying higher moments and i.i.d. respectively; $\hat{M}_1^d(\hat{p}_0), \hat{M}_0^d(\hat{p}_0)$, finite sample-corrected generalized spectral tests derived under time-varying higher moments and i.i.d. respectively; $LY_1(\bar{p}), LY_2(\bar{p})$, Li and Yu's (1994) test derived assuming var(ε_i) = 1 known and without assuming it; the Bartlett kernel is used for $\hat{M}_1(\hat{p}_0), \hat{M}_0(\hat{p}_0), \hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$. DGP S.1: $Y_t = \psi_i\varepsilon_i$, $\psi_i = 0.15 + 0.8\psi_{i-1} + 0.05Y_{i-1}^2$, $\varepsilon_i = z_i$, $z_i \sim$ i.i.d. exp (1); DGP S.2: $Y_t = \psi_i\varepsilon_i$, $\psi_i = 0.15 + 0.8\psi_{i-1} + 0.05Y_{i-1}^2$, $\varepsilon_i = \exp(\sqrt{h_i}z_i)/\exp(\frac{1}{2}h_i)$, $h_i = 0.5 + 0.5\varepsilon_{i-1}^2$, $z_i \sim$ i.i.d. exp (1).
DGP indicates data-generating process and ACD autoregressive conditional duration.

0 in probability rather slowly. Therefore, $\hat{\theta}$ may significantly distort the sizes of $\hat{M}_1(p)$ and $\hat{M}_0(p)$ in finite samples, as is observed in the simulation study that follows.

In practice, one could use a bootstrap procedure to approximate the finite sample distributions of $\hat{M}_1(p)$ and $\hat{M}_0(p)$. A naive bootstrap could be used for $\hat{M}_0(p)$ when $\{\varepsilon_i\}$ is i.i.d. and this is expected to yield accurate sizes in finite samples. However, even the naive bootstrap is computationally costly because, to account for the impact of parameter estimation uncertainty in finite samples, it will involve reestimation of the null ACD model using bootstrap samples. When the null ACD model is nonlinear, estimation can be rather involved. Moreover, the naive bootstrap cannot be used when $\{\varepsilon_i\}$ is not i.i.d. Sophisticated bootstraps (e.g. block bootstraps) are needed to take into account unknown serial dependence in higher-order conditional moments in $\{\varepsilon_i\}$. Here, we will use a convenient finite sample correction that can purge the impact of parameter estimation uncertainty of the test statistics and the resulting test statistics still follow the convenient null asymptotic N(0, 1) distribution. This is achieved by adopting Wooldridge's (1990a) device which has also been used in Hong and Lee (2007) for tests of time series regression models with additive errors.

5.1. Wooldridge's device

Wooldridge (1990a, 1990b, 1991) proposes a novel approach to robust, moment-based parametric specification testing for possibly dynamic time-series models. Specifically, Wooldridge (1990a) considers the null hypothesis

$$E[e_i(\theta_0)|I_{i-1}] = 0 \quad \text{for some } \theta_0 \in \Theta,$$

where $e_i(\theta)$ is a measurable, possibly vector-valued function. In the present context, $e_i(\theta) = \mathbf{i}[\varepsilon_i(\theta) - 1]$. Wooldridge (1990a) uses a weighting function $A_i(\theta) \in I_{i-1}$ and checks if $E[A_i(\theta_0)e_i(\theta_0)] = 0$ by using the sample moment

$$\hat{m} \equiv \frac{1}{n}\sum_{i=1}^{n} m_i(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \hat{A}_i\hat{e}_i,$$

where $m_i(\theta) = A_i(\theta)e_i(\theta)$, $\hat{A}_i = A_i(\hat{\theta}), \hat{e}_i = e_i(\hat{\theta})$ and $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator of θ_0. Straightforward algebra shows that

$$\sqrt{n}\hat{m} = n^{-1/2}\sum_{i=1}^{n}[m_i(\theta_0) + A_i(\theta_0)\Phi_i(\theta_0)(\hat{\theta} - \theta_0)] + O_P(n^{-1/2}),$$

where

$$\Phi_i(\theta_0) \equiv E\left[\frac{\partial}{\partial\theta}e_i(\theta_0)\,|\,I_{i-1}\right].$$

Thus, the asymptotic distribution of $\sqrt{n}\hat{m}$ is jointly determined by $n^{-1/2}\sum_{i=1}^{n} m_i(\theta_0)$ and $\sqrt{n}(\hat{\theta}-\theta_0)$, unless the expected derivative $\Phi_i(\theta_0)=0$ under $\mathbb{H}_0$. Here, the impact of $\sqrt{n}(\hat{\theta}-\theta_0)$ is because of the sampling variation of parameter estimation.

To remove the impact of parameter estimation uncertainty of $\hat{\theta}$ on the asymptotic distribution of $\sqrt{n}\hat{m}$, Wooldridge (1990a) first purges from $\hat{A}_i$ its linear projection onto $\hat{\Phi}_i$, a consistent estimator of $\Phi_i(\theta_0)$, and then considers the modified sample moment

$$\hat{m}_d \equiv \frac{1}{n}\sum_{i=1}^{n} m_i^d(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{A}_i - \hat{\Phi}_i'\hat{\beta})\hat{e}_i,$$

where $m_i^d(\theta) = [A_i(\theta) - \Phi_i(\theta)'\beta(\theta)]e_i(\theta)$, $\beta(\theta) = \{E[\Phi_i(\beta)\Phi_i(\beta)']\}^{-1}E[\Phi_i(\beta)A_i(\beta)]$, $\hat{\beta}$ is the ordinary least square (OLS) estimator of regressing $\hat{A}_i$ on $\hat{\Phi}_i$. It can be shown that for any $\sqrt{n}$-consistent estimator $\hat{\theta}$,

$$\sqrt{n}\hat{m}_d = n^{-1/2}\sum_{i=1}^{n} m_i^d(\theta_0) + O_P(n^{-1/2}).$$

Thus, the asymptotic distribution of $\sqrt{n}\hat{m}_d$ is robust to parameter estimation uncertainty because it is not affected by any $\sqrt{n}$-consistent estimator $\hat{\theta}$ up to $O_P(n^{-1/2})$. In other words, the asymptotic distribution of $\sqrt{n}\hat{m}_d$ remains unchanged when $\hat{\theta}$ is replaced with θ_0. An asymptotic chi-squared test can be obtained by forming a quadratic form in $\sqrt{n}\hat{m}_d$. It may be emphasized that Wooldridge's (1990a) device does not imply that $\sqrt{n}\hat{m}_d$ has a better asymptotic approximation than $\sqrt{n}\hat{m}$ in finite samples, or vice versa. However, it generates a new set of moment conditions $\sqrt{n}\hat{m}_d$ that is robust to parameter estimation uncertainty up to $O_P(n^{-1/2})$. Consequently, its asymptotic distribution does not depend on any $\sqrt{n}$ -consistent estimator $\hat{\theta}$. This makes the test based on $\sqrt{n}\hat{m}_d$ rather convenient.

Although Wooldridge's (1990a) device may not deliver a better asymptotic distribution approximation for a test based on $\sqrt{n}\hat{m}_d$, it ideally suits our purpose of improving the finite sample performance of the generalized spectral derivative tests $\hat{M}_1(p)$ and $\hat{M}_0(p)$ in (10) and (11). Intuitively, with a new set of moment conditions, it can make the asymptotically negligible higher-order terms in $\hat{M}_1(p)$ and $\hat{M}_0(p)$ that are associated with $\hat{\theta}$ vanish faster to 0, thus yielding better sizes in finite samples. Next, we first describe how Wooldridge's (1990a) device can be adopted to $\hat{M}_1(p)$ and $\hat{M}_0(p)$ and then explain the rationale behind the improvement of the asymptotic normal approximation for $\hat{M}_1(p)$ and $\hat{M}_0(p)$.

Although $\hat{M}_1(p)$ and $\hat{M}_0(p)$ are more complicated than Wooldridge's (1990a) test statistic, Wooldridge's (1990a) device can be applied to each generalized covariance derivative $\hat{\sigma}_j^{(1,0)}(0,v)$, which has a similar structure to $\hat{m}$, with $\hat{e}_i = \mathbf{i}(\hat{\varepsilon}_i - 1)$ and $\hat{A}_i = \hat{\phi}_{i-|j|}(v)$. The analysis here is more involved because of the need to integrate out the parameter v. Put $\phi_i(v) = e^{\mathbf{i}v\varepsilon_i} - \varphi(v)$ and $\eta_j(v) = E[G_i\phi_{i-j}(v)]$ for $j > 0$ where $G_i = \frac{\partial}{\partial\theta}\ln\psi(I_{i-1},\theta_0)$ as in Assumption A.5, and let $\tilde{\sigma}_j^{(1,0)}(0,v)$ be defined in the same way as $\hat{\sigma}_j^{(1,0)}(0,v)$ with $\{\varepsilon_i\}_{i=1}^n$ replacing $\{\hat{\varepsilon}_i\}_{i=1}^n$. Then, by a Taylor series expansion around θ_0, we have for each given $j > 0$,

$$\hat{\sigma}_j^{(1,0)}(0,v) = \tilde{\sigma}_j^{(1,0)}(0,v) - \mathbf{i}\eta_j(v)'(\hat{\theta}-\theta_0) + O_P[(n-j)^{-1}].$$

For most ACD models, where $\psi(I_{i-1},\theta_0)$ is a function of lagged dependent variables $\{Y_{i-j}\}$ and lagged expected durations $\{\psi_{i-j}\}$, $\eta_j(v)$ is non-zero at least for some $j > 0$. Consequently, for each given j, the asymptotic distribution of $\hat{\sigma}_j^{(1,0)}(0,v)$ is jointly determined by $\tilde{\sigma}_j^{(1,0)}(0,v)$ and $\eta_j(v)'(\hat{\theta}-\theta_0)$. The asymptotic distribution of $\hat{M}_1(p)$ or $\hat{M}_0(p)$, however, is only determined by the terms $\{\tilde{\sigma}_j^{(1,0)}(0,v)\}_{j=1}^{n-1}$, because $\hat{M}_1(p)$ or $\hat{M}_0(p)$ is a cumulative weighted sum of $\int|\hat{\sigma}_j^{(1,0)}(0,v)|^2\,\mathrm{d}W(v)$ over many lags, and the cumulative effect of the terms $\{\eta_j(v)'(\hat{\theta}-\theta_0)\}_{j=1}^{n-1}$ is of smaller order of magnitude than that for the terms $\{\tilde{\sigma}_j^{(1,0)}(0,v)\}_{j=1}^{n-1}$ given $\sum_{j=0}^{\infty}||\eta_j(v)|| < \infty$, which in turn is implied by the α-mixing condition on $\{\varepsilon_i, G_i'\}'$ in Assumption A.5.

Although the terms $\{\eta_j(v)'(\hat{\theta}-\theta_0)\}_{j=1}^{n-1}$ together are asymptotically negligible for the asymptotic distribution of $\hat{M}_1(p)$ and $\hat{M}_0(p)$, they vanish to 0 slowly and thus may affect the finite sample distributions of $\hat{M}_1(p)$ and $\hat{M}_0(p)$. This is indeed the case, as revealed in the simulation study in Section 7. To alleviate this, we introduce a modified sample generalized covariance function

$$\hat{\gamma}_j^{(1,0)}(0,v) = (n-|j|)^{-1}\sum_{i=|j|+1}^{n}\mathbf{i}(\hat{\varepsilon}_i-1)\hat{h}_{i-|j|}(v), \quad j=0,\pm1,\ldots,\pm(n-1), \tag{12}$$

where $\hat{h}_{i-|j|}(v)$ is the OLS residual of regression $\hat{\phi}_{i-j}(v)$ on the log-gradient vector $\hat{G}_i = \frac{\partial}{\partial\theta}\ln\psi(I_{i-1}^{\dagger},\hat{\theta})$; that is, $\hat{h}_{i-|j|}(v) = \hat{\phi}_{i-|j|}(v) - \hat{G}_i'\hat{\beta}_{|j|}(v)$, where

$$\hat{\beta}_{|j|}(v) = \left(\sum_{i=1}^{n}\hat{G}_i\hat{G}_i'\right)^{-1}\sum_{i=|j|+1}^{n}\hat{G}_i\hat{\phi}_{i-|j|}(v). \tag{13}$$

Following an analogous reasoning to Wooldridge (1990a), we have that for each $j > 0$,

$$\hat{\gamma}_j^{(1,0)}(0,v) = \tilde{\gamma}_j^{(1,0)}(0,v) + O_P[(n-j)^{-1/2}],$$

where

$$\tilde{\gamma}_j^{(1,0)}(0,v) = \frac{1}{n-j}\sum_{i=j+1}^{n} \mathbf{i}(\varepsilon_i - 1)h_{i-j}(v),$$

$h_{i-j}(v) = \phi_{i-j}(v) - G_i'\beta_j(v)$, and $\beta_j(v) = [E(G_iG_i)]^{-1}E[G_i\phi_{i-j}(v)]$. In other words, the impact of parameter estimation uncertainty has been effectively purged of $\hat{\gamma}_j^{(1,0)}(0,v)$. One can thus expect that the tests based on $\hat{\gamma}_j^{(1,0)}(0,v)$ will perform better than the tests that are based on $\hat{\sigma}_j^{(1,0)}(0,v)$ in finite samples, because the most slowly vanishing terms in $\hat{M}_1(p)$ and $\hat{M}_0(p)$ that are associated with parameter estimation uncertainty now disappear.

5.2. Finite sample corrected tests under m.d.s. innovations

When $\{\varepsilon_i - 1\}$ is m.d.s., we can obtain the modified test statistic:

$$\hat{M}_1^d(p) = \frac{\left[\sum_{j=1}^{n-1} k^2(j/p)(n-j)\int \left|\hat{\gamma}_j^{(1,0)}(0,v)\right|^2 \mathrm{d}W(v) - \hat{C}_1^d(p)\right]}{\sqrt{\hat{D}_1^d(p)}}, \tag{14}$$

where the centering and scaling factors

$$\hat{C}_1^d(p) = \sum_{j=1}^{n-1} k^2(j/p)\frac{1}{n-j}\sum_{i=j+1}^{n}(\hat{\varepsilon}_i^2 - 1)\int \left|\hat{h}_{i-j}(v)\right|^2 \mathrm{d}W(v),$$

$$\hat{D}_1^d(p) = 2\sum_{j=1}^{n-2}\sum_{l=1}^{n-2} k^2\left(\frac{j}{p}\right)k^2\left(\frac{l}{p}\right)\int\int \left|\frac{1}{n-\max(j,l)}\sum_{i=\max(j,l)+1}^{n}(\hat{\varepsilon}_i^2 - 1)\hat{h}_{i-j}(v)\hat{h}_{i-l}(v')\right|^2 \mathrm{d}W(v)\,\mathrm{d}W(v').$$

We expect a finite sample improvement of the asymptotic normal approximation for $\hat{M}_1^d(p)$, because its asymptotically negligible higher-order terms vanish to 0 in probability faster than the higher-order terms in $\hat{M}_1(p)$. This is confirmed in our simulation study. The finite sample improvement is achieved by combining Wooldridge's device and our non-parametric testing approach. As noted earlier, Wooldridge's device alone does not necessarily improve the finite sample performance. Intuitively, for each $\hat{\sigma}_j^{(1,0)}(0,v)$, there is an impact of parameter estimation uncertainty. The Taylor series expansion of $(n-j)^{1/2}\hat{\sigma}_j^{(1,0)}(0,v)$ around θ_0 reveals that replacing $\hat{\theta}$ for θ_0 affects the asymptotic distribution of $(n-j)^{1/2}\hat{\sigma}_j^{(1,0)}(0,v)$. Although the cumulative effect of replacing $\hat{\theta}$ for θ_0 becomes asymptotically negligible for $\hat{M}_1(p)$ when we use an increasing number of lags, the sampling variation of $\hat{\theta}$ may still significantly affect the finite sample distribution of $\hat{M}_1(p)$. In contrast, a Taylor series expansion of $(n-j)^{1/2}\hat{\gamma}_j^{(1,0)}(0,v)$ around θ_0 reveals that the asymptotic distribution of $(n-j)^{1/2}\hat{\gamma}_j^{(1,0)}(0,v)$ is the same as that of $(n-j)^{1/2}\tilde{\gamma}_j^{(1,0)}(0,v)$. By using $\hat{\gamma}_j^{(1,0)}(0,v)$, we can effectively reduce the impact of parameter estimation uncertainty to a higher order for each lag order j. As a result, robustness of $\hat{M}_1^d(p)$ to parameter estimation uncertainty in finite samples is achieved.

5.3. Finite sample-corrected tests under i.i.d. innovations

The finite sample correction is also applicable to $\hat{M}_0(p)$ when $\{\varepsilon_i\}$ is i.i.d. In this case, the modified test statistic is:

$$\hat{M}_0^d(p) = \frac{\left[\sum_{j=1}^{n-1} k^2(j/p)(n-j)\int |\hat{\gamma}_j^{(1,0)}(0,v)|^2 \mathrm{d}W(v) - \hat{C}_0^d(p)\right]}{\sqrt{\hat{D}_0^d(p)}}, \tag{15}$$

where the centering and scaling factors

$$\hat{C}_0^d(p) = (\hat{s}^2 - 1)\int |\hat{h}(v)|^2 \mathrm{d}W(v)\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right),$$

$$\hat{D}_0^d(p) = 2(\hat{s}^2 - 1)^2\sum_{j=1}^{n-2}\sum_{l=1}^{n-2} k^2(\frac{j}{p})k^2(l/p)\int\int \left|\frac{1}{n-\max(j,l)}\sum_{i=\max(j,l)+1}^{n}\hat{h}_{i-j}(v)\hat{h}_{i-l}(v')\right|^2 \mathrm{d}W(v)\,\mathrm{d}W(v'),$$

where, as before, $\hat{s}^2 = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^2$ and $\hat{h}(v) = n^{-1}\sum_{i=1}^n \hat{h}_i(v)$. We note that the asymptotic variance estimator $\hat{D}_0^d(p)$ under the i.i.d. case is more complicated than the asymptotic variance estimator $\hat{D}_0(p)$ for the original test $\hat{M}_0(p)$. This is because $\{h_{i-j}(v)\}$ is not i.i.d. even when $\{\phi_{i-j}(v)\}$ is i.i.d.

5.4. Asymptotic distributions of finite sample-corrected tests

To derive the null limiting distribution of $\hat{M}_1^d(p)$ and $\hat{M}_0^d(p)$, we impose the following additional regularity conditions:

Assumption A.9. *$E[\frac{\partial}{\partial\theta}\ln\psi(\mathrm{I}_{i-1},\theta)\frac{\partial}{\partial\theta'}\ln\psi(\mathrm{I}_{i-1},\theta)]$ is non-singular for all $\theta \in \Theta$.*

Assumption A.10. *Let $I_i^\dagger$ be an observed information set available at time t_i that may contain some assumed initial values. Then*

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left\{E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}\ln\,\psi(I_{i-1}^\dagger,\theta)-\frac{\partial}{\partial\theta}\ln\psi(I_{i-1},\theta)\right\|\right]^2\right\}^{1/2}\leqslant C.$$

We derive the asymptotic distributions of $\hat{M}_1^d(p)$ and $\hat{M}_0^d(p)$ under $\mathbb{H}_0$.

Theorem 2. *Suppose Assumptions A.1–A.10 hold, and $p=cn^\lambda$ for $0<\lambda<(3+\frac{1}{4b-2})^{-1}$ and $0<c<\infty$. Then (i) $\hat{M}_1^d(p)-\hat{M}_1(p)\longrightarrow^{p}0$, and $\hat{M}_1^d(p)\longrightarrow^{d}N(0,1)$ under $\mathbb{H}_0$. (ii) Suppose in addition $\{\varepsilon_i\}$ is i.i.d. Then $\hat{M}_0^d(p)-\hat{M}_0(p)\longrightarrow^{d}0$, and $\hat{M}_0^d(p)\longrightarrow^{d}N(0,1)$.*

Like the original tests $\hat{M}_1(p)$ and $\hat{M}_0(p)$ in (10) and (11), we obtain the convenient null asymptotic N(0, 1) distribution for the finite sample-corrected tests $\hat{M}_1^d(p)$ and $\hat{M}_0(p)$. Indeed Theorem 2 implies that $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ are asymptotically equivalent under $\mathbb{H}_0$. However, $\hat{M}_1^d(p)$ is expected to have a better finite sample performance, as is confirmed in the simulation study that follows.

The asymptotic equivalence between $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ under $\mathbb{H}_0$ has an important implication. Although we do not formally analyse it, we expect that the asymptotic equivalence between $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ will continue to hold under a suitable class of local alternatives to $\mathbb{H}_0$ (see Wooldridge, 1990a, p. 29, for a similar discussion on the original and finite sample-corrected parametric *m*-tests). In other words, $\hat{M}_1^d(p)$ will be asymptotically as powerful as $\hat{M}_1(p)$ under a class of local alternatives. This implies that the finite sample correction improves sizes in finite samples and does not suffer from asymptotic power loss.

We summarize the procedures to implement the modified tests $\hat{M}_1^d(p)$ and $\hat{M}_0^d(p)$:

1. *Step 1*: Obtain a $\sqrt{n}$-consistent estimator $\hat{\theta}$ [e.g. QMLE in (7)] for the ACD model $\psi(I_{i-1},\theta)$, and save the estimated standardized residual $\hat{\varepsilon}_i=Y_i/\psi(I_{i-1}^\dagger,\hat{\theta})$.
2. *Step 2*: Compute the log-gradient vector $\hat{G}_i=\frac{\partial}{\partial\theta}\ln\psi(I_{i-1}^\dagger,\hat{\theta})$. The calculation of $\hat{G}_i$ is convenient for most commonly used ACD models in practice. Although $\hat{G}_i$ may have a tedious closed form expression, it usually satisfies some simple recursive relationship, which can be used to calculate $\hat{G}_i$ recursively with some assumed initial values.
3. *Step 3*: For each lag order j from 1 to $n-1$, run an OLS regression of $\hat{\phi}_{i-j}(v)=e^{\mathbf{i}v\hat{\varepsilon}_{i-j}}-\hat{\varphi}(v)$ on $\hat{G}_i$, where $\hat{\varphi}(v)=n^{-1}\sum_{i=1}^{n}e^{\mathbf{i}v\hat{\varepsilon}_i}$ and we set $\hat{\phi}_i(v)=0$ for $i\le 0$. (Alternatively, one could use the OLS estimator $\hat{\beta}_j(v)=(\sum_{i=j+1}^{T}\hat{G}_i\hat{G}_i')^{-1}\sum_{i=j+1}^{n}\hat{G}_i\hat{\phi}_{i-j}(v)$ for $0<j<T$.) Save the estimated residual $\hat{h}_{i-j}(v)$. If the kernel $k(\cdot)$ has a bounded support (i.e. $k(z)=0$ if $|z|>1$), then it suffices to run regressions for j from 1 to p.
4. *Step 4*: Compute the finite sample-corrected test statistic $\hat{M}_1^d(p)$ in (14) or $\hat{M}_0^d(p)$ in (15).
5. *Step 5*: Compare $\hat{M}_1^d(p)$ or $\hat{M}_0^d(p)$ with an upper-tailed N(0, 1) critical value (e.g. 1.65 at the 5% level), and reject $\mathbb{H}_0$ at a given level if $\hat{M}_1^d(p)$ or $\hat{M}_0^d(p)$ is larger than the critical value.

6. ASYMPTOTIC POWER

We now investigate the asymptotic power of the proposed tests, particularly the impact of the finite sample correction on the power of the tests under $\mathbb{H}_A$. For this purpose, we define the covariance function

$$\gamma_j(u,v)=\mathrm{cov}[e^{\mathbf{i}u(\varepsilon_i-1)},h_{i-j}(v)],\quad u,v\in\mathbb{R}\text{ and }j>0,\tag{16}$$

where $h_{i-j}(v)=\phi_{i-j}(v)-G_i'\beta_j(v)$, and $\beta_j(v)=[E(G_iG_i')]^{-1}E[G_i\phi_{i-j}(v)]$. We can state Theorem 3 below, the main result of this section.

Theorem 3. *Suppose Assumptions A.1–A.7 hold, and $p=cn^\lambda$ for $0<\lambda<1/2$ and $0<c<\infty$. Then*

(i)

$$(p^{1/2}/n)\hat{M}_1(p)\xrightarrow{p}\left[2D\int_0^\infty k^4(z)\,\mathrm{d}z\right]^{-1/2}\sum_{j=1}^{\infty}\int\left|\sigma_j^{(1,0)}(0,v)\right|^2\mathrm{d}W(v),$$

where

$$D=\sigma^4\sum_{j=-\infty}^{\infty}\int\int|\sigma_j(u,v)|^2\,\mathrm{d}W(u)\,\mathrm{d}W(v),\quad\sigma^2=E(\varepsilon_1^2)-1,$$

and

$$(p^{1/2}/n)\hat{M}_0(p)\xrightarrow{p}\left[2D_0\int_0^\infty k^4(z)\,\mathrm{d}z\right]^{-1/2}\sum_{j=1}^{\infty}\int\left|\sigma_j^{(1,0)}(0,v)\right|^2\mathrm{d}W(v),$$

where $D_0 \equiv \sigma^4 \int\int |\sigma_0(u,v)|^2 \, dW(u) \, dW(v)$.

(ii) *Suppose in addition Assumptions A.9 and A.10 hold. Then*

$$(p^{1/2}/n)\hat{M}_1^d(p) \xrightarrow{p} \left[2D_d \int_0^\infty k^4(z)\,dz\right]^{-1/2} \sum_{j=1}^\infty \int \left|\gamma_j^{(1,0)}(0,v)\right|^2 dW(v),$$

where

$$D_d = \sigma^4 \sum_{j=-\infty}^\infty \int\int |\rho_j(u,v)|^2 \, dW(u)\, dW(v), \quad \rho_j(u,v) = \mathrm{cov}\left[h_i(u), h_{i-j}(v)\right].$$

Moreover,

$$(p^{1/2}/n)\hat{M}_0^d(p) \xrightarrow{p} \left[2D_d \int_0^\infty k^4(z)\,dz\right]^{-1/2} \sum_{j=1}^\infty \int \left|\gamma_j^{(1,0)}(0,v)\right|^2 dW(v).$$

The constant D, in Theorem 3(i) takes into account the impact of serial dependence in conditioning functions $\{\phi_{i-j}(v), j > 0\}$, which generally exists even under $\mathbb{H}_0$, because of the presence of serial dependence in the conditional dispersion and higher-order conditional moments of $\{\varepsilon_i\}$. As a result, $D > D_0$ when $\{\varepsilon_i\}$ is not i.i.d. This implies that $\hat{M}_0(p)$ will tend to have a large value than $\hat{M}_1(p)$ for sample sizes sufficiently large when $\{\varepsilon_i\}$ is not i.i.d. under $\mathbb{H}_A$. Thus, $\hat{M}_0(p)$ is expected to be more powerful than $\hat{M}_1(p)$ under $\mathbb{H}_A$ when both tests are applicable to test $\mathbb{H}_0$. Of course, the finite sample performance might tell a different story.

The constant D_d takes into account the impact of the finite sample correction. It depends on the serial dependence in $\{h_{i-j}(v), j > 0\}$, which exists even when $\{\varepsilon_i\}$ is i.i.d., because $\beta_j(v)$ is generally non-zero for most ACD models. Both the modified tests $\hat{M}_1^d(p)$ and $\hat{M}_0^d(p)$ are asymptotically equally powerful because they converge to the same probability limit.

As was noted in Section 2, one could test $\mathbb{H}_0$ by using the additive error $\xi_i = Y_i - \psi(I_{i-1},\theta) = \varepsilon_i[\psi_i^0 - \psi(I_{i-1},\theta)]$ rather than the standardized error $\varepsilon_i = Y_i/\psi(I_{i-1},\theta)$. However, $\{\xi_i\}$ is conditionally heteroskedastic under $\mathbb{H}_0$ even when $\{\varepsilon_i\}$ is i.i.d. Therefore, a test based on $\{\xi_i\}$ is expected to be asymptotically less powerful than a test based on $\{\varepsilon_i\}$, because $\{\xi_i\}$ displays 'more' serial dependence in higher-order moments than $\{\varepsilon_i\}$. To some extent, this is similar in spirit to the relative efficiency gain of the generalized least squares estimator over the ordinary least squares estimator when there exists conditional heteroskedasticity. Moreover, the use of $\{\varepsilon_i\}$ rather than $\{\xi_i\}$ allows weaker moment conditions on the DGP. In particular, an integrated ACD(1,1) model which is strictly but not weakly stationary is allowed. In this case, $\{\varepsilon_i\}$ is still weakly stationary but $\{\xi_i\}$ is not.

Because $\beta_j(v)$ is generally non-zero for (non-Markovian) ACD models, we have $D \neq D_d$ and $\sigma_j^{(1,0)}(0,v) \neq \gamma_j^{(1,0)}(0,v)$. As a result, $\hat{M}_1^d(p)$ and $\hat{M}_1(p)$ are not asymptotically equivalent under $\mathbb{H}_A$ in terms of Bahadur's (1960) asymptotic slope criterion because they do not converge to the same probability limit after multiplied by the rate $p^{1/2}/n$. Unlike the case under $\mathbb{H}_0$, where the auxiliary regressions have no impact on the asymptotic distribution of $\hat{M}_1^d(p)$, the auxiliary regressions have impact on the probability limit of $(p^{1/2}/n)\hat{M}_1^d(p)$ under $\mathbb{H}_A$ and thus affect the asymptotic efficiency of the test in terms of Bahadur's (1960) criterion.

To investigate how the auxiliary regressions may affect the asymptotic power of $\hat{M}_1(p)$, we assume that the autoregression function $E(\varepsilon_i|\varepsilon_{i-j}) \neq 1$ at some lag $j > 0$ under $\mathbb{H}_A$. Then we have $\int |\sigma_j^{(1,0)}(0,v)|^2 \, dW(v) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on $\mathbb{R}$. It follows from Theorem 3 that $P[\hat{M}_1(p) > C(n)] \to 1$ for any sequence of constants $C(n) = o(n/p^{1/2})$, and so the original test $\hat{M}_1(p)$ has asymptotic unit power at any given significance level $\alpha \in (0,1)$. Thus, $\hat{M}_1(p)$ has omnibus power against a wide variety of linear and nonlinear ACD alternatives with unknown lag structure. It avoids the blindness of searching for different alternatives when one has no prior information.

Theorem 3 also indicates that the power of the modified test $\hat{M}_1^d(p)$ depends on whether $\gamma_j^{(1,0)}(0,v) \neq 0$ at least for some $j > 0$ under $\mathbb{H}_A$. Generally we have $\gamma_j^{(1,0)}(0,v) \neq \sigma_j^{(1,0)}(0,v)$ under $\mathbb{H}_A$. However, if we have either (i) $E[G_i(\varepsilon_i - 1)] = 0$ or (ii) $\beta_j(v) = 0$ for all $j > 0$ under $\mathbb{H}_A$, then $\gamma_j^{(1,0)}(0,v) = \sigma_j^{(1,0)}(0,v)$ for all $v \in \mathbb{R}$. In these cases, $\hat{M}_1^d(p)$ has the same consistency property (i.e. asymptotic unit power) as $\hat{M}_1(p)$, although their probability limits may be different, because of the fact that $D \neq D_d$ generally. As noted earlier, we generally have $\beta_j(v) \neq 0$ for some $j > 0$ for non-Markovian ACD models. However, Case (i) that $E[G_i(\varepsilon_i - 1)] = 0$ can arise under $\mathbb{H}_A$ even when $\psi(I_{i-1},\theta)$ contains lagged dependent variables and lagged innovations. In particular, when $\hat{\theta}$ is the QMLE in (7), then $\theta^* = p\lim \hat{\theta}$ will satisfy the- first order condition that

$$E[G_i(\varepsilon_i - 1)] = E[\frac{\partial}{\partial\theta}\ln\psi(I_{i-1},\theta^*)\varepsilon_i] = 0$$

even under $\mathbb{H}_A$.

When $E[G_i(\varepsilon_i - 1)] \neq 0$ and $\beta_j(v) \neq 0$ at least for some $j > 0$, we generally have $\gamma_j^{(1,0)}(0,v) \neq 0$ if $\sigma_j^{(1,0)}(0,v) \neq 0$, although $\sigma_j^{(1,0)}(0,v) \neq \gamma_j^{(1,0)}(0,v)$. In this case, $\hat{M}_1^d(p)$ has the same consistency property as $\hat{M}_1(p)$. However, there exists certain model misspecification against which the modified test $\hat{M}_1^d(p)$ has no power. This arises when $\gamma_j^{(1,0)}(0,v) = 0$ for all $j > 0$ but $\sigma_j^{(1,0)}(0,v) \neq 0$ for some $j > 0$. Let

$$\alpha \equiv [E(G_i G_i')]^{-1} E[G_i(\varepsilon_i - 1)]$$

be the least squares coefficient of regressing $\varepsilon_i - 1$ on the log-gradient vector G_i. Then the possibility that $\gamma_j^{(1,0)}(0,v) = 0$ for all $j > 0$ but $\sigma_j^{(1,0)}(0,v) \neq 0$ for all $j > 0$ can arise if and only if $\alpha \neq 0$ and

$$\sigma_j^{(1,0)}(0,v) \equiv \mathrm{cov}[\mathbf{i}(\varepsilon_i - 1), e^{\mathbf{i}v\varepsilon_{i-j}}] = \mathbf{i}\alpha' E\left(G_i G_i'\right)\beta_j(v) = \mathbf{i}\mathrm{cov}[\alpha' G_i, G_i'\beta_j(v)] \quad \text{for all } j > 0,$$

that is, if and only if the covariance between ε_i-1 and $e^{\mathbf{i}v\varepsilon_{i-j}}$ coincides with the covariance between their linear projections onto G_i. This occurs when the neglected dynamics takes the form of $E(\varepsilon_i|I_{i-1}) = 1 + \alpha'[G_i-E(G_i)]$ subject to the constraint that $\varepsilon_i \geq 0$, The finite sample-corrected test $\hat{M}_1^d(p)$ has no power against this (pathological) misspecification. This is the price that we have to pay when using Wooldridge's (1990a) device. However, we emphasize that the gain in the size improvement from using Wooldridge's device for our tests overwhelms the possible power loss in detecting misspecification in the direction of the log gradient G_i. More importantly, if the QMLE in (7), which is always available, is used so that $E[G_i(\varepsilon_i-1)] = 0$, $\hat{M}_1^d(p)$ will be able to detect such pathological misspecification, and achieve the same consistency property as the original test $\hat{M}_1(p)$.

Because of using a relatively large number of lag orders, the proposed tests have power against misspecification at higher-order lags. In particular, they are expected to have good power against long memory ACD models (Jasiak, 1998). At the same time, they do not suffer from the loss of a large number of degrees of freedom, thanks to the use of $k^2(\cdot)$. Most non-uniform kernels discount higher-order lags. This enhances good power against the alternatives whose serial dependence in mean decays to 0 when lag order j increases, as shown in the simulation study. Thus, our tests can check a large number of lags without losing too many degrees of freedom. This feature is not shared by chi-squared-type tests with a large number of lags, which essentially give equal weighting to each lag. Equal weighting is not fully efficient when a large number of lags is used.

Once the model $\psi(I_{i-1},\theta)$ is rejected by $\hat{M}_1^d(p)$, one may like to further explore possible sources of misspecifications in an ACD model. For this purpose, we can construct a sequence of tests similar in spirit to $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ by using the partial derivatives with respect to v:

$$\sigma_j^{(1,l)}(0,0) \equiv \frac{\partial^l}{\partial v^l}\sigma_j^{(1,0)}(0,v)|_{v=0} = \mathrm{cov}[\mathbf{i}(\varepsilon_i - 1), (\mathbf{i}\varepsilon_{i-j})^l],$$
$$\gamma_j^{(1,l)}(0,0) \equiv \frac{\partial^l}{\partial v^l}\gamma_j^{(1,0)}(0,v)|_{v=0} = \mathrm{cov}\left[\mathbf{i}(\varepsilon_i - 1), (\mathbf{i}\varepsilon_{i-j})^l - G_i'\frac{d^l}{dv^l}\beta_j(0)\right], \quad l = 0,1,2,\ldots.$$

For l = 1, 2, 3, 4, tests based on these derivatives can check whether there exists linear correlation, dispersion-in-mean, skewness-in-mean and kurtosis-in-mean effects respectively. These derivative tests may reveal valuable information about the nature of model misspecification. For space, we do not derive the concrete form of these diagnostic test statistics.

7. MONTE CARLO EVIDENCE

We now investigate the finite sample performance of the proposed tests $\hat{M}_1(p), \hat{M}_0(p), \hat{M}_1^d(p)$ and $\hat{M}_0^d(p)$ in comparison with the most closely related modified portmanteau test of Li and Yu (2003).

7.1. Simulation design

7.1.1. Size

To examine the sizes of the tests under $\mathbb{H}_0$, we consider an ACD(1,1) DGP

$$\begin{cases} Y_i = \psi_i^0\varepsilon_i, \\ \psi_i^0 = \alpha_0 + \beta_0\psi_{i-1}^0 + \gamma_0 Y_{i-1}, \end{cases}$$

where (i) $\{\varepsilon_i\} \sim$ i.i.d. exp (1), or (ii) $\varepsilon_i = \exp(\sqrt{h_i}z_i)/\exp(\frac{1}{2}h_i)$, $h_i = 0.5 + 0.5\varepsilon_{i-1}^2$, where $\{z_i\} \sim$ i.i.d. N(0,1). Under (i), $\{\varepsilon_i\}$ is i.i.d. with $E(\varepsilon_i) = 1$, whereas under (ii), $E(\varepsilon_i \mid I_{i-1}) = 1$ but its conditional variance var$(\varepsilon_i \mid I_{i-1}) = h_i$ follows an autoregressive conditional heteroskedasticity (ARCH)-type behaviour. We set $\theta_0 = (\alpha_0, \beta_0, \gamma_0)' = (0.15, 0.8, 0.05)'$.

The null model for Y_i is an ACD(1,1) specification:

$$\psi_i = \alpha + \beta\psi_{i-1} + \gamma Y_{i-1}, \tag{17}$$

where $\psi_i = \psi(I_{i-1},\theta)$ and $\theta = (\alpha,\beta,\gamma)'$. We estimate θ using QMLE in (7). The standardized model error $\{\varepsilon_i(\theta_0)\}$ is i.i.d. under (i), and displays dispersion clustering under (ii). This allows us to examine the robustness of the tests to time-varying conditional dispersion. We consider three sample sizes: n = 500, 1000 and 2000. For each n, we generate 1000 data sets using the GAUSS Windows version 5.0 random number generator on a PC. For each iteration, we first generate n + 500 observations and then discard the first 500 to reduce the impact of some assumed initial values, $\bar{Y}_0$ and $\bar{\psi}_0$.

To compute the generalized spectral derivative tests, we use the N(0,1) CDF truncated on [−3,3] for the weighting function $W(\cdot)$, and use the Parzen kernel

$$k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3 & \text{if } |z| \leq 1/2, \\ 2(1-|z|)^3 & \text{if } 1/2 \leq |z| < 1, \\ 0 & \text{otherwise,} \end{cases}$$

which has a bounded support and is computationally efficient. For the choice of lag order p, we use a data-driven lag order $\hat{p}_0$ via the plug-in method considered in Hong (1999), with the Bartlett kernel $\bar{k}(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$ used in a preliminary generalized spectral derivative estimator. To certain extent, the data-driven lag order $\hat{p}_0$ lets the data speak for an appropriate lag order, but it still involves the choice of the preliminary bandwidth $\bar{p}$ which is somewhat arbitrary. To examine the impact of the choice of the preliminary bandwidth $\bar{p}$, we consider $\bar{p} = 5, 10, 20$ respectively.

We also consider Li and Yu's (2005) test. Let $\hat{r} = [\hat{r}(1), \hat{r}(2), \ldots, \hat{r}(p)]'$, where $\hat{r}(j)$ is the sample autocorrelation function of the estimated standardized model error $\{\hat{\varepsilon}_i\}$. Li and Yu's (2005) test statistic is

$$\mathrm{LY}(p) = n\hat{r}'(I - X\hat{H}^{-1}X')^{-1}\hat{r} \xrightarrow{d} \chi^2_p,$$

where X is a $p \times d$ matrix with the jth row $X_j = n^{-1}\sum_{i=j+1}^{n} \hat{G}_i\hat{\varepsilon}_i(\hat{\varepsilon}_{i-j} - 1)$, and $\hat{H}$ is the negative Hessian matrix

$$\hat{H} = \frac{1}{n}\sum_{i=1}^{n}(1 - 2\hat{\varepsilon}_i)\hat{G}_i\hat{G}_i' + \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i\frac{\partial^2 \ln \psi(I_{i-1}^{\dagger}, \hat{\theta})}{\partial\theta\partial\theta'}.$$

We consider two versions of Li and Yu's (2005) test. The first is considered in Li and Yu (2003), where the variance estimator is set to unity, under the assumption that $\{\varepsilon_i\} \sim$ i.i.d. exp (1); the second version is based on the sample variance. We note that we could use some simpler expressions for X and $\hat{H}$ by exploiting the fact $E(\varepsilon_i \mid I_{i-1}) = 1$ under $\mathbb{H}_0$. However, the resulting test statistics often lead to negative values so they were not used in the simulation study.

7.1.2. Power

To examine the power of the tests for dynamic misspecification (i.e. lag order misspecification) and neglected nonlinearity in $\psi(I_{i-1},\theta)$, we consider the following DGPs:

DGP P.1 [ACD(2, 1)]:

$$\psi_i^0 = 0.15 + 0.10Y_{i-1} + 0.05Y_{i-2} + 0.80\psi_{i-1}^0;$$

DGP P.2 [log-ACD(1, 1)]:

$$\ln \psi_i^0 = 0.1334 + 0.115 \ln Y_{i-1} + 0.7749 \ln \psi_{i-1};$$

DGP P.3 [exponential ACD]:

$$\ln \psi_i^0 = -0.0806 + 0.2061\varepsilon_{i-1} - 0.1309|\varepsilon_{i-1} - 1| + 0.9149 \ln \psi_{i-1}^0;$$

DGP P.4 [threshold ACD(1, 1)]:

$$\psi_i^0 = \begin{cases} 0.020 + 0.257Y_{i-1} + 0.847\psi_{i-1}^0 & \text{if } Y_{i-1} \leq 3.79, \\ 1.808 + 0.027Y_{i-1} + 0.501\psi_{i-1}^0 & \text{if } Y_{i-1} > 3.79, \end{cases}$$

where $\{\varepsilon_i\} \sim$ i.i.d. exp (1).

DGP P.1 is an ACD(2, 1) process used in Meitz and Teräsvirta (2006). This allows us to investigate dynamic misspecification (i.e. lag order misspecification) of an ACD(1, 1) model. DGP P.2–P.4 are some popular nonlinear ACD processes. DGP P.2, a log-ACD model, introduced by Bauwen and Giot (2000), is more flexible than a linear ACD model because no restrictions are required on the sign of its coefficients. It allows for nonlinear effects of short and long durations, without requiring the estimation of additional parameters. DGP P.3, an exponential ACD model, is introduced by Dufour and Engle (2000) in a similar spirit to Nelson's (1991) exponential generalized ARCH model. This allows for a piecewise linear news impact function kinked at the mean $E(\varepsilon_i) = 1$. DGP P.4, a threshold ACD model, proposed by Zhang *et al.* (2001), is a simple but powerful approach to allow subregimes to achieve different persistences in ψ_i^0, which allows for greater flexibility compared with the ACD models.

By construction, the proposed tests have power when $\psi(I_{i-1},\theta)$ is misspecified for ψ_i^0. Thus, they are expected to have power against all DGPs P.1–P.4 when an ACD(1, 1) model in (17) is used. On the other hand, they also have power when an ACD(1, 1) model is correctly specified for ψ_i^0 but the parameter estimator $\hat{\theta}$ is not consistent for θ_0 where θ_0 is defined in $\mathbb{H}_0$. We note that parameter choices are also relevant under the alternative. In this simulation study, our main purpose is to examine whether the proposed tests have reasonable power in distinguishing specification differences, namely in distinguishing an ACD(1, 1) model from other alternatives. We have used the QMLE $\hat{\theta}$ which is consistent under $\mathbb{H}_0$.

For each of the DGPs P.1–P.4, we consider three sample sizes: $n = 500$, 1000 and 2000. For each n, we generate 500 data sets under each DGP. For each data set, we generate $n + 500$ observations and then discard the first 500 to reduce the impact of the choice of some initial values.

7.2. Monte Carlo evidence

Table 1 reports the empirical rejection rates of the tests under $\mathbb{H}_0$ at the 10% and 5% significance levels, using the asymptotic theory. We first consider the robust tests $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$. No matter whether $\{\varepsilon_i\}$ is i.i.d. or not under $\mathbb{H}_0$, $\hat{M}_1(\hat{p}_0)$ underrejects $\mathbb{H}_0$ substantially at both the 10% and 5% levels, even when $n = 2000$. In contrast, the finite sample-corrected test $\hat{M}_1^d(\hat{p}_0)$ has reasonable sizes in most cases, whether $\{\varepsilon_i\}$ is i.i.d. or not. These results are consistent with our theory that parameter estimation uncertainty may have non-trivial impact on $\hat{M}_1(\hat{p}_0)$ in finite samples, and the finite sample correction gives better asymptotic approximation. They indicate the relative robustness of the finite sample-corrected test $\hat{M}_1^d(\hat{p}_0)$ to parameter estimation uncertainty, illustrating the merits of adopting Wooldridge's (1990a) device to the generalized spectral tests for ACD models. Intuitively, parameter estimation is like a calibration that makes the demeaned estimated standardized model residuals $\{\hat{\varepsilon}_i - 1\}$ look more like an m.d.s., leading to underrejection of the original test $\hat{M}_1(\hat{p}_0)$. However, Wooldridge's (1990a) device effectively removes the impact of parameter estimation, and as a result, $\hat{M}_1(\hat{p})$ has better sizes.

When $\{\varepsilon_i\}$ is i.i.d., $\hat{M}_0(\hat{p}_0)$ has slightly better sizes than $\hat{M}_1(\hat{p}_0)$ in many cases, apparently because of the fact that $\hat{M}_0(\hat{p}_0)$ exploits the implication of the i.i.d. assumption for $\{\varepsilon_i\}$. However, it still underrejects $\mathbb{H}_0$, because of the impact of parameter estimation uncertainty in finite samples. On the other hand, when $\{\varepsilon_i\}$ is not i.i.d., $\hat{M}_0(\hat{p}_0)$ shows rather strong overrejection, because of the fact that the asymptotic variance estimator $\hat{D}_0(p)$ underestimates the true asymptotic variance D when $\{\varepsilon_i\}$ is not i.i.d. (recall $D > D_0$ in this case where the expressions of D and D_0 are given in Theorem 3). This highlights the importance of taking into account time-varying conditional dispersion and higher-order conditional moments in $\{\varepsilon_i\}$. Failure to do so will cause strong overrejection for the generalized spectral derivative tests.

When $\{\varepsilon_i\}$ is i.i.d. exp (1), both versions $\text{LY}_1(\bar{p})$ and $\text{LY}_2(\bar{p})$ of Li and Yu's (2005) test have reasonable sizes in most cases for all three sample sizes. When $\{\varepsilon_i\}$ is not i.i.d., $\text{LY}_1(\bar{p})$, which assumes $\text{var}(\varepsilon_i) = 1$, strongly overrejects $\mathbb{H}_0$. However, $\text{LY}_2(\bar{p})$, which uses the sample variance estimator, is still reasonable in many cases, although it tends to overreject $\mathbb{H}_0$ in some cases (particularly for $n = 2000$). This is interesting because $\text{LY}_2(\bar{p})$ is not asymptotically valid under case (ii), because of the fact that the parameter estimator $\hat{\theta}$ is not MLE in this case. A robust version of $\text{LY}_2(\bar{p})$ when $\hat{\theta}$ is not MLE could be developed along the reasoning of Li and Yu (2005), but it is not yet available in the literature.

Next, we turn to the power of the tests, reported in Tables 2 and 3. Table 2 reports the empirical rejection rates of the tests at the 10% and 5% levels under DGPs P.1 and P.2 using empirical critical values, which provide a fair ground to compare different tests. We find that $\hat{M}_1^d(\hat{p}_0)$ has a similar power to $\hat{M}_0^d(\hat{p}_0)$. This is consistent with the asymptotic theory that they are asymptotically equally powerful under $\mathbb{H}_A$. The $\text{LY}_1(\bar{p})$ and $\text{LY}_2(\bar{p})$ tests also have very similar power. The $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ tests have better power than $\text{LY}_1(\bar{p})$ and $\text{LY}_2(\bar{p})$ in most cases, particularly when lag order $\bar{p}$ is large. This highlights the merit of discounting higher-order lags via the squared kernel function $k^2(\cdot)$, and the use of the plug-in data-driven method to select $\hat{p}_0$. The power of $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ is relatively robust to the choice of the preliminary lag order $\bar{p}$.

For the original tests, $\hat{M}_1(\hat{p}_0)$ has slightly better power than $\hat{M}_0(\hat{p}_0)$ in some cases. Both $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_0(\hat{p}_0)$ have a little higher power than $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ respectively. There is some loss of power because of the use of the finite sample correction under DGP P.1.

Table 2. Empirical powers of tests

	$M_1(\hat{p}_0)$		$M_0(\hat{p}_0)$		$M_1^d(\hat{p}_0)$		$M_0^d(\hat{p}_0)$		$\text{LY}_1(\bar{p})$		$\text{LY}_2(\bar{p})$	
$\bar{p}$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DGP P.1: ACD(2,1)												
$N = 500$												
5	22.4	14.6	20.8	13.8	17.4	11.8	17.4	10.8	19.2	12.2	17.8	10.8
10	22.8	14.4	20.2	13.0	16.2	11.8	16.6	10.4	19.0	11.6	19.2	12.2
20	18.8	11.0	18.4	10.8	17.0	10.2	16.6	11.4	19.4	13.0	18.0	12.4
$N = 1000$												
5	26.8	17.8	25.6	16.8	20.2	13.4	20.8	13.4	27.2	16.0	25.2	13.8
10	27.0	16.8	25.4	15.8	20.6	13.6	20.4	12.8	20.4	10.6	20.8	9.6
20	23.2	13.0	21.4	11.6	20.6	11.0	20.2	10.0	17.6	11.8	18.8	11.0
$N = 2000$												
5	38.2	28.4	37.4	24.8	32.6	22.4	32.0	21.4	35.0	23.0	35.2	25.0
10	38.2	28.6	37.2	25.0	31.8	22.6	32.0	21.6	30.4	18.6	29.2	18.2
20	31.8	22.6	31.8	20.2	28.6	18.8	29.8	16.8	26.2	15.4	26.0	15.2
DGP P.2: log-ACD												
$N = 500$												
5	27.2	15.8	18.0	10.2	28.4	18.0	20.2	11.4	13.6	6.2	11.0	5.2
10	26.6	15.4	18.0	9.6	28.8	18.4	20.0	11.2	14.2	8.0	10.0	7.0
20	23.4	12.0	13.8	6.8	25.2	16.4	17.4	9.0	17.6	7.6	11.2	5.6
$N = 1000$												
5	46.4	30.4	36.6	22.8	47.0	35.4	38.0	38.0	22.0	13.0	20.4	8.4
10	45.8	29.8	35.8	21.6	45.8	35.4	37.8	37.8	21.4	12.4	16.2	8.6
20	40.4	26.6	29.4	18.4	42.8	28.6	33.2	33.2	23.6	16.4	18.6	10.4
$N = 2000$												
5	76.2	66.2	69.2	54.0	78.6	66.8	72.0	58.0	44.6	30.0	39.8	21.2
10	76.2	66.2	69.2	54.0	78.4	66.8	72.0	58.0	46.0	31.4	32.4	17.8
20	74.8	58.6	64.6	46.2	77.4	62.8	71.2	51.6	45.2	31.6	31.8	20.0

500 iterations; $\hat{M}_1(\hat{p}_0), \hat{M}_0(\hat{p}_0)$, generalized spectral tests derived under time-varying higher moments and i.i.d. respectively; $\hat{M}_1^d(\hat{p}_0), \hat{M}_0^d(\hat{p}_0)$, finite sample-corrected generalized spectral tests derived under time-varying higher moments and i.i.d. respectively; $LY_1(\bar{p}), LY_2(\bar{p})$, Li and Yu's (1994) test derived assuming $\text{var}(\varepsilon_i) = 1$ known and without assuming it; the Bartlett kernel is used for $\hat{M}_1(\hat{p}_0), \hat{M}_0(\hat{p}_0), \hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$. DGP P.1, $\psi_i = 0.15 + 0.10Y_{i-1} + 0.05Y_{i-2} + 0.80\psi_{i-1}$, where $\{\varepsilon_i\} \sim$ i.i.d. exp (1); DGP P.2, $\ln\psi_i = 0.1334 + 0.115\ln Y_{i-1} + 0.7749\ln\psi_{i-1}$, where $\{\varepsilon_i\} \sim$ i.i.d. exp (1).
DGP indicates data-generating process and ACD autoregressive conditional duration.

Table 3. Empirical powers of tests

	$\hat{M}_1(\hat{p}_0)$		$\hat{M}_0(\hat{p}_0)$		$\hat{M}_1^d(\hat{p}_0)$		$\hat{M}_0^d(\hat{p}_0)$		$\text{LY}_1(\bar{p})$		$\text{LY}_2(\bar{p})$	
$\bar{p}$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DGP P.3: EACD(1,1)												
$N = 500$												
5	19.6	12.0	14.8	7.8	20.6	13.8	15.0	9.0	13.6	6.4	11.8	6.8
10	19.6	12.6	14.4	7.6	21.0	13.8	14.8	7.8	14.4	9.6	12.6	7.6
20	16.8	10.0	12.0	6.4	20.0	11.0	13.6	8.2	14.4	7.2	14.4	7.6
$N = 1000$												
5	33.4	19.0	27.4	15.2	33.0	22.6	26.6	15.6	18.8	10.6	18.0	7.6
10	33.6	18.6	26.2	13.8	33.2	22.0	26.8	15.2	16.4	8.8	13.4	6.2
20	30.4	17.6	21.8	12.8	30.0	20.2	25.2	13.2	18.0	11.8	16.2	10.2
$N = 2000$												
5	58.6	43.6	50.4	32.6	59.6	45.6	52.2	36.6	28.8	16.8	26.6	15.8
10	58.6	43.6	50.4	32.8	59.2	45.6	52.0	36.6	28.0	18.8	25.2	13.0
20	53.2	37.2	43.0	27.8	55.8	39.4	48.2	32.2	29.8	18.8	24.8	12.8
DGP P.4: 2-TACD												
$N = 500$												
5	18.6	11.4	14.2	7.6	20.2	14.2	17.2	10.2	9.6	3.2	9.0	4.4
10	18.4	11.8	15.2	7.0	21.2	14.2	17.4	10.0	9.6	3.8	8.6	4.4
20	17.0	8.6	12.0	5.8	20.4	11.2	16.0	9.4	9.6	5.0	9.6	3.8
$N = 1000$												
5	29.4	18.2	24.2	13.8	29.8	20.4	24.8	17.4	12.4	6.6	9.8	5.0
10	29.4	18.4	23.6	13.4	29.6	20.6	24.6	16.6	14.2	7.6	12.8	5.6
20	27.8	16.0	21.2	13.0	26.8	16.4	24.6	13.8	12.6	6.8	11.0	5.2
$N = 2000$												
5	51.8	38.4	45.6	28.6	52.6	41.6	49.2	34.2	16.0	10.4	15.0	9.6
10	51.4	38.4	45.6	29.0	52.2	41.6	49.0	34.4	15.2	9.0	13.6	6.4
20	48.2	32.2	38.4	24.8	48.8	35.2	47.2	30.4	18.0	12.0	14.6	7.8

500 iterations; $\hat{M}_1(\hat{p}_0), \hat{M}_0(\hat{p}_0)$, generalized spectral tests derived under time-varying higher moments and i.i.d. respectively; $\hat{M}_1^d(\hat{p}_0), \hat{M}_0^d(\hat{p}_0)$, finite sample-corrected generalized spectral tests derived under time-varying higher moments and i.i.d. respectively; $LY_1(\bar{p}), LY_2(\bar{p})$, Li and Yu's (1994) test derived assuming $\text{var}(\varepsilon_i) = 1$ known and without assuming it; the Bartlett kernel is used for $\hat{M}_1(\hat{p}_0), \hat{M}_0(\hat{p}_0), \hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$. DGP P.3, $\ln\psi_i = -0.0806 + 0.2061\varepsilon_{i-1} - 0.1309|\varepsilon_{i-1} - 1| + 0.9149\ln\psi_{i-1}$, $\{\varepsilon_i\} \sim$ i.i.d. exp (1).

$$\text{DGP P.4, } \psi_i = \begin{cases} 0.020 + 0.257Y_{i-1} + 0.847\psi_{i-1} & \text{if } Y_{i-1} \leq 3.79, \\ 1.808 + 0.027Y_{i-1} + 0.501\psi_{i-1} & \text{if } Y_{i-1} > 3.79. \end{cases} \quad \{\varepsilon_i\} \sim \text{i.i.d. exp}(1).$$

DGP indicates data-generating process and TACD threshold autoregressive conditional duration.

Now, we consider DGP P.2 [log-ACD(1, 1)]. The $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$ tests have a little better power than $\hat{M}_0(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ respectively, whereas $\text{LY}_1(p)$ has slightly better power than $\text{LY}_2(p)$ in some cases. Unlike under DGP P.1 [ACD(2, 1)], the original tests $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_0(\hat{p}_0)$ have similar power to the modified tests $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ respectively. The tests $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ have substantially better power than $\text{LY}_1(p)$ and $\text{LY}_2(p)$, but the latter also have increasing power when the sample size n increases. The powers of all tests are robust to the choice of the lag order $\bar{p}$.

Table 3 reports the power patterns under DGP P.3 [EACD(1, 1)] and DGP P.4 [TACD(1, 1)]. They are similar to those under DGP P.2. In particular, the powers of the generalized spectral derivative tests for ACD models are substantially more powerful than Li and Yu's (2005) tests. These results highlight the merits of the generalized spectral tests in detecting nonlinear ACD alternatives from a linear ACD(1, 1) model.

In summary, we have observed the following stylized facts:

- The empirical sizes of the original generalized spectral derivative tests $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_0(\hat{p}_0)$ are substantially lower than the nominal significance levels, because of the impact of parameter estimation uncertainty in estimating the null ACD model. Wooldridge's (1990a) device can effectively reduce the impact of parameter estimation uncertainty – the empirical sizes of the finite sample-corrected tests $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(\hat{p}_0)$ are reasonable and robust to parameter estimation uncertainty in most cases, especially when the sample size becomes moderately large. The sizes of the generalized spectral tests are relatively robust to the choice of the preliminary lag order $\bar{p}$.
- The powers of the finite sample-corrected generalized spectral tests $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_0^d(p)$ are more or less similar to the original tests $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_0(\hat{p}_0)$ respectively for most cases (except for DGP P.1), suggesting that the finite sample correction does not suffer from power loss in detecting the alternatives under study. The robust tests $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_1(\hat{p}_0)$ are very slightly more powerful than $\hat{M}_0^d(\hat{p}_0)$ and $\hat{M}_0(\hat{p}_0)$ respectively in most cases.
- The power of the generalized spectral derivative tests is similar to the power of Li and Yu's tests in detecting dynamic (lag order) misspecification in a linear ACD alternative. As expected, the generalized spectral derivative tests have substantially better power than Li and Yu's tests against some popular nonlinear ACD models, but the latter also have some power when the sample size n becomes large.

8. CONCLUSION

Using a generalized spectral derivative approach, we develop a new class of specification tests for linear and nonlinear ACD models, where the dimension of the conditioning observable information set grows with time or is infinite, because of the non-Markovian property of most ACD models. The tests can detect a wide range of model misspecification in the conditional expected duration dynamics while being robust to time-varying conditional dispersion and other higher-order moments of unknown form in standardized innovations. The tests have a convenient null asymptotic N(0,1) distribution which is not affected by parameter estimation uncertainty. To remove the non-trivial impact of parameter estimation uncertainty in finite samples, we propose a finite sample correction by adopting Wooldridge's (1990a) device. This leads to reasonable size performances which are robust to parameter estimation uncertainty in finite samples. Like the original generalized spectral derivative tests, the finite sample-corrected tests can detect a wide range of model misspecification in the conditional expected duration while being robust to time-varying conditional dispersion and higher-order moments of unknown form in standardized innovations. Moreover, the consistency property of the original tests is preserved when QMLE is used to estimate the null ACD model. Both original and modified tests have reasonable power against a variety of dynamic misspecification and nonlinear ACD alternatives. These results indicate that the proposed tests with a finite sample correction can be a useful and reliable diagnostic tool in evaluating ACD models.

MATHEMATICAL APPENDIX

Proof of Theorem 1. For space, we only consider the proof of $\hat{M}_1(p)$; the proof for $\hat{M}_0(p)$ is similar and simpler. We first define a pseudo test statistic

$$\tilde{M}_1(p) = \frac{\left[\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \, \mathrm{d}W(v) - \tilde{C}_1(p)\right]}{\sqrt{\tilde{D}_1(p)}},$$

where

$$\tilde{\sigma}_j^{(1,0)}(0,v) = n_j^{-1} \sum_{i=j+1}^{n} \mathbf{i}(\varepsilon_i - 1)\phi_{i-j}(v), \phi_i(v) = e^{iv\varepsilon_i} - \varphi(v), \varphi(v) = E(e^{iv\varepsilon_i}),$$

and

$$\tilde{C}_1(p) = \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \int n_j^{-1} \sum_{i=j+1}^{n} (\varepsilon_i^2 - 1)\left|\phi_{i-j}(v)\right|^2 \mathrm{d}W(v),$$

$$\tilde{D}_1(p) = \sum_{j=1}^{n-2}\sum_{l=1}^{n-2} k^2\left(\frac{j}{p}\right) k^2\left(\frac{l}{p}\right) \int\int \left|\frac{1}{n-\max(j,l)} \sum_{i=\max(j,l)+1}^{n} (\varepsilon_i^2 - 1)\phi_{i-j}(v)\phi_{i-l}(v)\right|^2 \mathrm{d}W(u)\,\mathrm{d}W(v).$$

It suffices to show Theorems A.1–A.3 next. Theorem A.1 implies that the use of the estimated standardized model residual $\{\hat{\varepsilon}_i\}_{i=1}^n$ rather than the unobserved $\{\varepsilon_i\}_{i=1}^n$ has no impact on the limit distribution of $\hat{M}_1(p)$. Theorem A.2 implies that the use of the truncated disturbances $\{\varepsilon_{q,i}\}_{i=1}^n$ rather than $\{\varepsilon_i\}_{i=1}^n$ has no impact on the limit distribution of $\hat{M}_1(p)$ when q is sufficiently large. The assumption that $\varepsilon_{q,i}$ is independent of $\{\varepsilon_{i-j}\}_{j=q+1}^{\infty}$ when q is large simplifies a great deal the proof of asymptotic normality of $\hat{M}_1(p)$. □

Theorem A.1. *Under the conditions of Theorem* 1, $\hat{M}_1(p) - \tilde{M}_1(p) \to^p 0$.

Theorem A.2. *Let* $\tilde{M}_{q,1}(p)$ *be defined as* $\tilde{M}_1(p)$ *with* $\{\varepsilon_{q,i}\}_{i=1}^n$ *replacing* $\{\varepsilon_i\}_{i=1}^n$, *where* $\{\varepsilon_{q,i}\}$ *is as in Assumption A.8. Then under the conditions of Theorem* 1 *and* $q = p^{(1+(1/4b-2))(ln^2 n)^{1/(2b-1)}}$, $\tilde{M}_{q,1}(p) - \tilde{M}_1(p) \to^p 0$.

Theorem A.3. *Under the conditions of Theorem* 1 *and* $q = p^{(1+(1/4b-2))(\ln^2 n)^{\frac{1}{2b-1}}}$, $\tilde{M}_{1q}(p) \to_d N(0,1)$.

Proof of Theorem A1. Noting that $\varepsilon_i(\theta) \equiv Y_i/\psi(I_{i-1},\theta)$ in (2), where I_{i-1} is the information set from period i to the infinite past, we can write

$$\hat{\varepsilon}_i \equiv \frac{Y_i}{\psi(I_{i-1}^{\dagger},\hat{\theta})} = \varepsilon_i(\hat{\theta})\frac{\psi(I_{i-1},\hat{\theta}) - \psi(I_{i-1}^{\dagger},\hat{\theta})}{\psi(I_{i-1}^{\dagger},\hat{\theta})} + \varepsilon_i(\hat{\theta}). \tag{A1}$$

It follows from (A1) and Markov's inequality that

$$\sum_{i=1}^{n}[\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})]^2 = \sum_{i=1}^{n}\varepsilon_i^2(\hat{\theta})\left[\frac{\psi(I_{i-1},\hat{\theta}) - \psi(I_{i-1}^{\dagger},\hat{\theta})}{\psi(I_{i-1}^{\dagger},\hat{\theta})}\right]^2 = O_P(1), \tag{A2}$$

where we have made use of the fact that

$$\begin{aligned} E\sum_{i=1}^{n}\varepsilon_i^2(\hat{\theta})\left[\frac{\psi(I_{i-1},\hat{\theta}) - \psi(I_{i-1}^{\dagger},\hat{\theta})}{\psi(I_{i-1}^{\dagger},\hat{\theta})}\right]^2 &\le \sum_{i=1}^{n}[E\sup_{\theta\in\Theta_0}\varepsilon_i^4(\theta)]^{1/2}\left\{E\sup_{\theta\in\Theta_0}\left[\frac{\psi(I_{i-1},\theta) - \psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right]^4\right\}^{1/2} \\ &\le C\sum_{i=1}^{n}\left\{E\sup_{\theta\in\Theta_0}\left[\frac{\psi(I_{i-1},\theta) - \psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right]^4\right\}^{1/2} \\ &\le C^2 \end{aligned}$$

by the Cauchy–Schwarz inequality and Assumptions A.2 and A.3.

Next, by the mean value theorem for $\varepsilon_i(\hat{\theta})$ around θ_0, we have

$$\varepsilon_i(\hat{\theta}) = \varepsilon_i(\theta_0) + \lambda_i(\bar{\theta})'(\hat{\theta} - \theta_0) = \varepsilon_i + \lambda_i(\bar{\theta})(\hat{\theta} - \theta_0) \tag{A3}$$

for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0, where

$$\lambda_i(\theta) \equiv \frac{\partial}{\partial\theta}\varepsilon_i(\theta) = -\varepsilon_i(\theta)\frac{\partial \ln\psi(I_{i-1},\theta)}{\partial\theta} = -\varepsilon_i(\theta)G_i(\theta),$$

with $G_i(\theta) \equiv (\partial/\partial\theta)\ln\psi(I_{i-1},\theta)$. It follows from (A2), the Cauchy–Schwarz inequality and Assumptions A.2–A.4 that

$$\sum_{i=1}^{n}[\varepsilon_i(\hat{\theta}) - \varepsilon_i]^2 \le \|\sqrt{n}(\hat{\theta} - \theta_0)\|^2 n^{-1}\sum_{i=1}^{n}\sup_{\theta\in\Theta_0}\|\lambda_i(\theta)\|^2 = O_P(1), \tag{A4}$$

where we made use of the facts that $E\sup_{\theta\in\Theta_0}||\lambda_i(\theta)||^2 \le [E\sup_{\theta\in\Theta_0}||\lambda_i(\theta)||^4]^{1/2}$, and

$$E\sup_{\theta\in\Theta_0}\|\lambda_i(\theta)\|^{\max(2\nu,4)} = E\sup_{\theta\in\Theta_0}\|\varepsilon_i(\theta)G_i(\theta)\|^{\max(2\nu,4)} \le [E\sup_{\theta\in\Theta_0}\varepsilon_i^{\max(4\nu,8)}(\theta)]^{1/2}[E\sup_{\theta\in\Theta_0}\|G_i(\theta)\|^{\max(4\nu,8)}]^{1/2} \le C \tag{A5}$$

given Assumption A.2, where $\nu > 1$ is in Assumption A.2. Both (A2) and (A4) imply

$$\sum_{i=1}^{n}(\hat{\varepsilon}_i - \varepsilon_i)^2 = O_P(1). \tag{A6}$$

Now we put $n_j = n - |j|$. Observe that $p \to \infty$, $p/n \to 0$, $p^{-1}\sum_{j=1}^{n-1}k^r(j/p) \to \int_0^\infty k^r(z)\,dz$ for $r = 2,4$ given Assumption A.6. To show $\hat{M}_1(p) - \tilde{M}_1(p) \to^p 0$, it suffices to show that $p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)] = O_P(n^{-1/2})$, $p^{-1}[\hat{D}_1(p) - \tilde{D}_1(p)] \to^p 0$, $\tilde{D}_1(p) = pD\int_0^\infty k^4(z)\,dz[1 + o(1)]$ for some bounded constant D, and

$$p^{-1/2}\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j[|\hat{\sigma}_j^{(1,0)}(0,v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0,v)|^2]\,dW(v) \xrightarrow{p} 0. \tag{A7}$$

For space, we focus on the proof of (A7); the proofs for $p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)] = O_P(n^{-1/2})$, $p^{-1}[\hat{D}_1(p) - \tilde{D}_1(p)] \xrightarrow{p} 0$, and $\tilde{D}_1(p) = p\int_0^\infty k^4(z)\,dz[1 + o(1)]$ are relatively straightforward. We note that the convergence rate $O_P(n^{-1/2})$ for $p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)]$ implies that replacing $\hat{C}_1(p)$ with $\tilde{C}_1(p)$ has asymptotically negligible impact given $p/n \to 0$.

To show (A7), we decompose

$$\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j\left[|\hat{\sigma}_j^{(1,0)}(0,v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0,v)|^2\right]dW(v) = \hat{A}_1 + 2\,\mathrm{Re}(\hat{A}_2),$$

where

$$\hat{A}_1 = \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \left|\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v)\right|^2 \mathrm{d}W(v), \tag{A8}$$

$$\hat{A}_2 = \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j [\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v)] \tilde{\sigma}_j^{(1,0)}(0,v)^* \mathrm{d}W(v), \tag{A9}$$

where $\mathrm{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\tilde{\sigma}_j^{(1,0)}(0,v)^*$ is the complex conjugate of $\tilde{\sigma}_j^{(1,0)}(0,v)$. Then, (A7) follows from Propositions A.1 and A.2 below, and $p \to \infty$ as $n \to \infty$. □

Proposition A.1. *Under the conditions of Theorem 1,* $\hat{A}_1 = O_P(1)$ *and* $p^{-1/2}\hat{A}_1 \to^p 0$.

Proposition A.2. *Under the conditions of Theorem 1,* $p^{-1/2}\hat{A}_2 \to^p 0$.

Proof of Proposition A.1. Throughout, we put $\hat{\delta}_i(v) \equiv e^{\mathbf{i}v\hat{\varepsilon}_i} - e^{\mathbf{i}v\varepsilon_i}$. Then straightforward algebra yields that for $j > 0$,

$$\begin{aligned}
&\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v) \\
&= \mathbf{i}n_j^{-1} \sum_{i=j+1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)\hat{\delta}_{i-j}(v) + \mathbf{i}n_j^{-1} \sum_{i=j+1}^{n} (\varepsilon_i - 1)\hat{\delta}_{i-j}(v) \\
&\quad + \mathbf{i}n_j^{-1} \sum_{i=j+1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)\phi_{i-j}(v) - \mathbf{i}[\hat{\varphi}(v) - \varphi(v)]n_j^{-1} \sum_{i=j+1}^{n} (\hat{\varepsilon}_i - \varepsilon_i) \\
&\quad - \mathbf{i}[\hat{\varphi}(v) - \varphi(v)]n_j^{-1} \sum_{i=j+1}^{n} (\varepsilon_i - 1) \\
&= \mathbf{i}\left[\hat{B}_{1j}(v) + \hat{B}_{2j}(v) + \hat{B}_{3j}(v) - \hat{B}_{4j}(v) - \hat{B}_{5j}(v)\right], \text{ say.}
\end{aligned} \tag{A10}$$

It follows that

$$\hat{A}_1 \le 8 \sum_{c=1}^{5} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{cj}(v)|^2 \mathrm{d}W(v).$$

Proposition A.1 follows from Lemmas A.1–A.5 below, and $p/n \to 0$.

Lemma A.1. $\sum_{j=1}^{n-1} k^2(j/p) n_j \int |\hat{B}_{1j}(v)|^2 \, dW(v) = O_P(p/n)$.

Lemma A.2. $\sum_{j=1}^{n-1} k^2(j/p) n_j \int |\hat{B}_{2j}(v)|^2 \, dW(v) = O_P(p/n)$.

Lemma A.3. $\sum_{j=1}^{n-1} k^2(j/p) n_j \int |\hat{B}_{3j}(v)|^2 \, dW(v) = O_P(1)$.

Lemma A.4. $\sum_{j=1}^{n-1} k^2(j/p) n_j \int |\hat{B}_{4j}(v)|^2 \, dW(v) = O_P(p/n)$.

Lemma A.5. $\sum_{j=1}^{n-1} k^2(j/p) n_j \int |\hat{B}_{5j}(v)|^2 \, dW(v) = O_P(p/n)$.

We now prove these lemmas sequentially. For notional convenience, we put $a_n(j) \equiv k^2(j/p)n_j^{-1}$.

Proof of Lemma A.1. By the Cauchy–Schwarz inequality and the inequality that $|e^{\mathbf{i}z_1} - e^{\mathbf{i}z_2}| \le |z_1 - z_2|$ for any real-valued variables z_1 and z_2, which implies $|\delta_i(v)| \le |v| \cdot |\hat{\varepsilon}_i - \varepsilon_i|$, we have

$$|\hat{B}_{1j}(v)|^2 \le \left[n_j^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_i-\varepsilon_i)^2\right]\left[n_j^{-1}\sum_{i=1}^{n}|\hat{\delta}_i(v)|^2\right] \le v^2\left[n_j^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_i-\varepsilon_i)^2\right]^2.$$

It follows from (A6), and Assumptions A.6 and A.7 that

$$\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j|\hat{B}_{1j}(v)|^2\,\mathrm{d}W \le \left[\sum_{j=1}^{n-1}a_n(j)\right]\left[\sum_{i=1}^{n}(\hat{\varepsilon}_i-\varepsilon_i)^2\right]^2\int v^2\,\mathrm{d}W(v) = O_P\left(\frac{p}{n}\right),$$

where we made use of the fact that

$$\sum_{j=1}^{n-1}a_n(j) = \sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j^{-1} = O\left(\frac{p}{n}\right) \tag{A11}$$

given $p = cn^{\lambda}$ for $\lambda \in (0, \frac{1}{2})$, as shown in Hong (1999, A.15, p. 1213). □

Proof of Lemma A.2. Using the inequality that $|e^{\mathbf{i}z} - 1 - \mathbf{i}z| \le |z|^2$ for any real-valued z, we have

$$|\hat{\delta}_i(v) - \mathbf{i}v(\hat{\varepsilon}_i-\varepsilon_i)e^{\mathbf{i}v\varepsilon_i}| \le v^2(\hat{\varepsilon}_i-\varepsilon_i)^2. \tag{A12}$$

Also, a second-order Taylor series expansion for $\varepsilon_i(\hat{\theta})$ around θ_0 yields

$$\varepsilon_i(\hat{\theta}) - \varepsilon_i = \lambda_i(\theta_0)'(\hat{\theta}-\theta_0) + \frac{1}{2}(\hat{\theta}-\theta_0)'\lambda_i'(\bar{\theta})(\hat{\theta}-\theta_0) \tag{A13}$$

for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0, where

$$\lambda_i'(\theta) \equiv \frac{\partial}{\partial\theta}\lambda_i(\theta) = -\varepsilon_i^2(\theta)G_i(\theta)G_i(\theta)' + \varepsilon_i(\theta)G_i'(\theta)$$

with $G_i'(\theta) \equiv (\partial/\partial\theta)G_i(\theta) = (\partial^2/\partial\theta\partial\theta')\ln\psi(I_{i-1},\theta)$. Then, (A12), (A13), $\hat{\varepsilon}_i - \varepsilon_i = [\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})] + [\varepsilon_i(\hat{\theta}) - \varepsilon_i]$ and $|e^{\mathbf{i}v\varepsilon_i}| = 1$ imply

$$|\hat{\delta}_i(v) - \mathbf{i}v\lambda_i(\theta_0)e^{\mathbf{i}v\varepsilon_{i-j}}(\hat{\theta}-\theta_0)| \le |v||\hat{\varepsilon}_i-\varepsilon_i(\hat{\theta})| + |v|\|\hat{\theta}-\theta_0\|^2\sup_{\theta\in\Theta}\|\lambda_i'(\theta)\| + v^2(\hat{\varepsilon}_i-\varepsilon_i)^2.$$

Therefore, by the definition of $\hat{B}_{2j}(v)$ in (A10), we obtain

$$\begin{aligned} n_j\big|\hat{B}_{2j}(v)\big| \le{}& |v|\sum_{i=j+1}^{n}|\varepsilon_i-1||\hat{\varepsilon}_{i-j}-\varepsilon_{i-j}(\hat{\theta})| + |v|\|\hat{\theta}-\theta_0\|\left\|\sum_{i=j+1}^{n}(\varepsilon_i-1)\lambda_{i-j}(\theta_0)e^{\mathbf{i}v\varepsilon_{i-j}}\right\| \\ &+ |v|\|\hat{\theta}-\theta_0\|^2\sum_{i=j+1}^{n}|\varepsilon_i-1|\sup_{\theta\in\Theta_0}\|\lambda_{i-j}'(\theta)\| + v^2\sum_{i=j+1}^{n}|\varepsilon_i-1|(\hat{\varepsilon}_{i-j}-\varepsilon_{i-j})^2. \end{aligned}$$

It follows that

$$\begin{aligned} &\sum_{j=1}^{n-1}k^2\int\left(\frac{j}{p}\right)n_j|\hat{B}_{2j}(v)|^2\,\mathrm{d}W \\ &\le 8\int v^2\,\mathrm{d}W(v)\sum_{j=1}^{n-1}a_n(j)\left[\sum_{i=1}^{n}|\varepsilon_i-1|\left|\hat{\varepsilon}_{i-j}-\varepsilon_{i-j}(\hat{\theta})\right|\right]^2 \\ &+ 8\left\|\sqrt{n}(\hat{\theta}-\theta_0)\right\|^2\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)\left\|n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_i-1)\xi_{i-j}(\theta_0)e^{\mathbf{i}v\varepsilon_{i-j}}\right\|^2\int v^2\,\mathrm{d}W(v) \\ &+ 8\left\|\sqrt{n}(\hat{\theta}-\theta_0)\right\|^4\left[n^{-1}\sum_{i=1}^{n}(\varepsilon_i-1)^2\right]\left[n^{-1}\sum_{i=1}^{n}\sup_{\theta\in\Theta_0}\left\|\lambda_i'(\theta)\right\|^2\right]\left[\sum_{j=1}^{n-1}a_n(j)\right]\int v^4\,\mathrm{d}W(v) \\ &+ 8\left\{\sum_{j=1}^{n-1}a_n(j)\left[\sum_{i=1}^{n}|\varepsilon_i-1|(\hat{\varepsilon}_{i-j}-\varepsilon_{i-j})^2\right]^2\right\}\int v^4\,\mathrm{d}W(v) \\ &= 8(\hat{F}_1+\hat{F}_2+\hat{F}_3+\hat{F}_4),\ \text{say}. \end{aligned} \tag{A14}$$

For the first term $\hat{F}_1$ in (A14), using Minkowski's inequality, the Cauchy–Schwarz inequality, (A1) and Assumptions A.2 and A.3, we have

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}|\varepsilon_i - 1|\cdot|\hat{\varepsilon}_{i-j} - \varepsilon_{i-j}(\hat{\theta})|\right]^2 &\leq \left[\sum_{i=1}^{n} E[(\varepsilon_i - 1)^2(\hat{\varepsilon}_{i-j} - \varepsilon_{i-j}(\hat{\theta}))^2]^{1/2}\right]^2\\
&\leq \left\{\sum_{i=1}^{n}[E\sup_{\theta\in\Theta_0}\varepsilon_i^8(\theta)]^{1/4}\left[E\sup_{\theta\in\Theta_0}\left|\frac{\psi(I_{i-1},\theta)-\psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right|^4\right]^{1/4}\right\}^2\\
&\leq C\sum_{i=1}^{n}\left[E\sup_{\theta\in\Theta_0}\left|\frac{\psi(I_{i-1},\theta)-\psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right|^4\right]^{1/4}\\
&\leq C^2.
\end{aligned}
$$

Hence, we have $\hat{F}_1 = O_P(p/n)$ by Markov's inequality, (A11) and Assumption A.7.
Next, for the second term $\hat{F}_2$ in (A14), using the fact that

$$
E\left\|\sum_{i=j+1}^{n}(\varepsilon_i - 1)\lambda_{i-j}(\theta_0)e^{\mathbf{i}v\varepsilon_{i-j}}\right\|^2 \leq Cn_j
$$

given the m.d.s. property of $\{\varepsilon_i - 1\}$ under $\mathbb{H}_0$ and (A5), we have $\hat{F}_2 = O_P(p/n)$ by Markov's inequality, (A11) and Assumptions A.4, A.6 and A.7.

For the third term $\hat{F}_3$ in (A14), we have $\hat{F}_3 = O_P(p/n)$ given Assumptions A.2, A.4–A.7, (A11) and the fact that

$$
E\sup_{\theta\in\Theta_0}\left\|\lambda_i'(\theta)\right\|^2 \leq 2[E\sup_{\theta\in\Theta}\varepsilon_i^8(\theta)E\sup_{\theta\in\Theta}\|G_i(\theta)\|^8]^{1/2} + 2[E\sup_{\theta\in\Theta}\varepsilon_i^4(\theta)E\sup_{\theta\in\Theta}\|G_i'(\theta)\|^4]^{1/2} \leq C \tag{A15}
$$

given Assumption A.2.

Finally, for the last term $\hat{F}_4$ in (A14), we have $\hat{F}_4 = O_P(p/n)$ because

$$
\begin{aligned}
\sum_{j=1}^{n-1} a_n(j)\left[\sum_{i=1}^{n}|\varepsilon_i - 1|(\hat{\varepsilon}_{i-j} - \varepsilon_{i-j})^2\right]^2 &\leq 2\sum_{j=1}^{n-1} a_n(j)\left[\sum_{i=1}^{n}|\varepsilon_i - 1|[\hat{\varepsilon}_{i-j} - \varepsilon_{i-j}(\hat{\theta})]^2\right]^2 + 2\sum_{j=1}^{n-1} a_n(j)\left[\sum_{i=1}^{n}|\varepsilon_i - 1|[\varepsilon_{i-j}(\hat{\theta}) - \varepsilon_{i-j}]^2\right]^2\\
&= O_P\left(\frac{p}{n}\right) + O_P\left(\frac{p}{n}\right),
\end{aligned} \tag{A16}
$$

where, for the first term in (A16), we have made use of the fact that, using (A1), Minkowski's inequality, the Cauchy–Schwarz inequality,

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}|\varepsilon_i - 1|[\hat{\varepsilon}_{i-j} - \varepsilon_{i-j}(\hat{\theta})]^2\right]^2 &\leq \left\{\sum_{i=1}^{n}[E(\varepsilon_i^8)E\sup_{\theta\in\Theta_0}\varepsilon_i^{16}(\theta)]^{1/8}\left[E\sup_{\theta\in\Theta_0}\left|\frac{\psi(I_{i-1},\theta)-\psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right|^8\right]^{1/4}\right\}^2\\
&\leq C\left\{\sum_{i=1}^{n}\left[E\sup_{\theta\in\Theta_0}\left|\frac{\psi(I_{i-1},\theta)-\psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right|^8\right]^{1/4}\right\}^2\\
&\leq C^2
\end{aligned}
$$

given Assumptions A.2; and for the second term in (A16), we have made use of the fact that, using the first-order Taylor series expansion for $\varepsilon_i(\hat{\theta})$ in (A3) and the Cauchy–Schwarz inequality,

$$
\left[\sum_{i=1}^{n}|\varepsilon_i - 1|[\varepsilon_{i-j}(\hat{\theta}) - \varepsilon_{i-j}]^2\right]^2 \leq \|\sqrt{n}(\hat{\theta} - \theta_0)\|^4\left[n^{-1}\sum_{i=1}^{n}(\varepsilon_i - 1)^2\right]\left[n^{-1}\sum_{i=1}^{n}\sup_{\theta\in\Theta}\|\lambda_i(\theta)\|^4\right] = O_P(1)
$$

given Assumptions A.2 and A.4, and (A5).

Because $\hat{F}_c = O_P(p/n)$ for $c = 1, 2, 3, 4$, we have from (A14) that

$$
\sum_{j=1}^{n-1}\int k^2\left(\frac{j}{p}\right)n_j|\hat{B}_{2j}(v)|^2\,\mathrm{d}W = O_P\left(\frac{p}{n}\right).
$$

This completes the proof of Lemma A.2. □

PROOF OF LEMMA A.3. Using that $\hat{\varepsilon}_i - \varepsilon_i = \hat{\varepsilon}_i - \varepsilon_i(\hat{\theta}) + \varepsilon_i(\hat{\theta}) - \varepsilon_i$, we first write

$$\hat{B}_{3j}(v) = n_j^{-1}\sum_{i=j+1}^{n}[\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})]\phi_{i-j}(v) + n_j^{-1}\sum_{i=j+1}^{n}[\varepsilon_i(\hat{\theta}) - \varepsilon_i]\phi_{i-j}(v) \equiv \hat{B}_{31j}(v) + \hat{B}_{32j}(v), \text{ say.} \tag{A17}$$

Given $|\phi_i(v)| \le 2$, (A11), and Assumptions A.2, A.3, A.6 and A.7, we have

$$\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{31j}(v)|^2 \, dW(v) \le 2\left[\sum_{i=1}^{n} |\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})|\right]^2 \sum_{j=1}^{n-1} a_n(j) \int dW(v) = O_P\left(\frac{p}{n}\right), \tag{A18}$$

where

$$\sum_{i=1}^{n} |\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})| = O_P(1) \tag{A19}$$

by Markov's inequality and the fact that, using (A1) and Assumption A.3,

$$\begin{aligned}
E\sum_{i=1}^{n} |\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})| &\le \sum_{i=1}^{n} [E \sup_{\theta\in\Theta} \varepsilon_i^2(\theta)]^{1/2} \left[E \sup_{\theta\in\Theta}\left|\frac{\psi(I_{i-1},\theta) - \psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right|^2\right]^{1/2} \\
&\le C\sum_{i=1}^{n}\left[E \sup_{\theta\in\Theta}\left|\frac{\psi(I_{i-1},\theta) - \psi(I_{i-1}^{\dagger},\theta)}{\psi(I_{i-1}^{\dagger},\theta)}\right|^8\right]^{1/8} \\
&\le C^2.
\end{aligned}$$

Next, using the second-order Taylor series expansion of $\varepsilon_i(\hat{\theta})$ in (A13), we have

$$\hat{B}_{32j}(v) = (\hat{\theta} - \theta_0)' n_j^{-1}\sum_{i=j+1}^{n} \lambda_i(\theta_0)\phi_{i-j}(v) + \frac{1}{2}(\hat{\theta} - \theta_0)'\left[n_j^{-1}\sum_{i=j+1}^{n} \lambda_i'(\bar{\theta})\phi_{i-j}(v)\right](\hat{\theta} - \theta_0),$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_0. We then have

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{32j}(v)|^2 \, dW(v) \\
&\le 2\left\|\sqrt{n}(\hat{\theta} - \theta_0)\right\|^2 \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right)\int \left\| n_j^{-1}\sum_{i=j+1}^{n} \lambda_i(\theta_0)\phi_{i-j}(v)\right\|^2 dW(v) \\
&+ 2\left\|\sqrt{n}(\hat{\theta} - \theta_0)\right\|^4 \left[n_j^{-1}\sum_{i=1}^{n} \sup_{\theta\in\Theta_0} \|\lambda_i'(\theta)\|^2\right]\left[\sum_{j=1}^{n-1} a_n(j)\right]\int dW(v).
\end{aligned} \tag{A20}$$

We now consider the first term in (A20). Put $\xi_j(v) \equiv E\{\lambda_i(\theta)\phi_{i-j}(v)\} = \text{cov}[\lambda_i(\theta), \phi_{i-j}(v)]$. Then, Assumptions A.2 and A.5 and a standard α-mixing inequality imply

$$\|\xi_j(v)\| \le [E\|\lambda_i(\theta)\|^{2v}]^{1/2v}[E|\phi_{i-j}(v)|^{2v}]^{1/2v}\alpha(j)^{(v-1)/v} \le C\alpha(j)^{(v-1)/v}. \tag{A21}$$

Moreover, given Assumptions A.4 and A.5, we have

$$E\left\| n_j^{-1}\sum_{i=j+1}^{n} \lambda_i(\theta_0)\phi_{i-j}(v) - \xi_j(v)\right\|^2 \le Cn_j^{-1}, \tag{A22}$$

using reasoning analogous to (A.7)–(A.10) in the proof of Thm 1 of Hong (1999, pp. 1212–1213). Consequently, from (A21), (A22), $|k(\cdot)| \le 1$, (A11) and $p/n \to 0$, we have

$$\begin{aligned}
\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) E\int \left\| n_j^{-1}\sum_{i=j+1}^{n} \lambda_i(\theta_0)\phi_{i-j}(v)\right\|^2 dW(v) &\le C\sum_{j=1}^{n-1}\int \|\xi_j(v)\|^2 \, dW(v) + C\sum_{j=1}^{n-1} a_n(j) \\
&= O(1) + O\left(\frac{p}{n}\right) = O(1).
\end{aligned} \tag{A23}$$

Next, for the second term in (A20), we have

$$\left\|\sqrt{n}(\hat{\theta}-\theta_0)\right\|^4\left[n_j^{-1}\sum_{i=1}^{n}\sup_{\theta\in\Theta_0}\|\lambda_i'(\theta)\|^2\right]\left[\sum_{j=1}^{n-1}a_n(j)\right]\int dW(v)=O_P(1)O_P(1)O_P\left(\frac{p}{n}\right)=O_P\left(\frac{p}{n}\right), \tag{A24}$$

by Assumption A.4, (A11), and (A15). The desired result follows from (A20), (A23) and (A24) as well as (A17) and (A18). □

Proof of Lemma A.4. *By the triangle inequality and the fact that* $|\hat{\delta}_i(v)| \leq |v|\cdot|\hat{\varepsilon}_i - \varepsilon_i|$, we have

$$\begin{aligned}|\hat{\varphi}(v)-\varphi(v)| &\leq \left|n^{-1}\sum_{i=1}^{n}\hat{\delta}_i(v)\right|+\left|n^{-1}\sum_{i=1}^{n}\phi_i(v)\right| \\ &\leq |v|n^{-1}\sum_{i=1}^{n}|\hat{\varepsilon}_i-\varepsilon_i|+\left|n^{-1}\sum_{i=1}^{n}\phi_i(v)\right|.\end{aligned} \tag{A25}$$

For the first term in (A25), we have

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}|\hat{\varepsilon}_i-\varepsilon_i| &\leq n^{-1}\sum_{i=1}^{n}|\hat{\varepsilon}_i-\varepsilon_i(\hat{\theta})|+n^{-1}\sum_{i=1}^{n}|\varepsilon_i(\hat{\theta})-\varepsilon_i| \\ &= O_P(n^{-1})+O_P(n^{-1/2})=O_P(n^{-1/2}),\end{aligned} \tag{A26}$$

given (A19), and the fact that

$$n^{-1}\sum_{i=1}^{n}|\varepsilon_i(\hat{\theta})-\varepsilon_i| \leq \|\hat{\theta}-\theta_0\|n^{-1}\sum_{i=1}^{n}\sup_{\theta\in\Theta_0}\|\lambda_i(\theta)\|=O_P(n^{-1/2})$$

by (A3), the Cauchy–Schwarz inequality, Assumption A.4 and (A5). Moreover, we have $\sup_{v\in\mathbb{R}}E\left|n^{-1}\sum_{i=1}^{n}\phi_i(v)\right|^2\leq Cn^{-1}$ using a standard α-mixing inequality that $|e[\phi_i(v)\phi_{i-j}(v)|\leq C\alpha(j)^{(v-1)/v}$. It follows from Markov's inequality that

$$\int\left|n^{-1}\sum_{i=1}^{n}\phi_i(v)\right|^2 dW(v)=O_P(n^{-1}). \tag{A27}$$

Combining (A25)–(A27), we have

$$\int|\hat{\varphi}(v)-\varphi(v)|^2 dW(v)=O_P(n^{-1}). \tag{A28}$$

It follows from (A6), (A11) and (A28) that

$$\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j\int|\hat{B}_{4j}(v)|^2 dW\leq n\int|\hat{\varphi}(v)-\varphi(v)|^2 dW(v)\left[\sum_{i=1}^{n}(\hat{\varepsilon}_i-\varepsilon_i)^2\right]\sum_{j=1}^{n-1}a_n(j)=O_P\left(\frac{p}{n}\right).$$

This completes the proof of Lemma A.4. □

Proof of Lemma A.5. Given (*A*11), (*A*28) and the m.d.s. property of $\{\varepsilon_i-1\}$ under $\mathbb{H}_0$, we have

$$\begin{aligned}\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j\int|\hat{B}_{5j}(v)|^2 dW(v) &\leq n\int|\hat{\varphi}(v)-\varphi(v)|^2 dW(v)\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)\left[n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_i-1)\right]^2 \\ &= O_P\left(\frac{p}{n}\right),\end{aligned}$$

where we made use of the fact that

$$E[n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_i-1)]^2\leq Cn_j^{-1}.$$ □

Proof of Proposition A.2. Given the decomposition in (A10), we have

$$\left|[\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v)]\tilde{\sigma}_j^{(1,0)}(0,v)^*\right| \le \sum_{c=1}^{5} |\hat{B}_{cj}(v)||\tilde{\sigma}_j^{(1,0)}(0,v)|, \tag{A29}$$

where $\{\hat{B}_{cj}(v)\}$ are defined in (A10). By the Cauchy–Schwarz inequality, we have for $c = 1, 2, 4, 5$,

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{cj}(v)||\tilde{\sigma}_j^{(1,0)}(0,v)|\, \mathrm{d}W(v) \\
&\le \left[\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{cj}(v)|^2\, \mathrm{d}W(v)\right]^{1/2} \left[\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\tilde{\sigma}_j^{(1,0)}(0,v)|^2\, \mathrm{d}W(v)\right]^{1/2} \\
&= O_P\left(\frac{p^{1/2}}{n^{1/2}}\right) O_P(p^{1/2}) = o_P(p^{1/2}),
\end{aligned} \tag{A30}$$

given Lemmas A.1, A.2, A.4 and A.5, and $p/n \to 0$, where

$$p^{-1} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\tilde{\sigma}_j^{(1,0)}(0,v)|^2\, \mathrm{d}W(v) = O_P(1)$$

by Markov's inequality, the m.d.s. hypothesis of $\{\varepsilon_i - 1\}$ under $\mathbb{H}_0$, and (A11).

It remains to consider the case with $c = 3$. By (A17) and the triangular inequality, we have

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{3j}(v)||\tilde{\sigma}_j^{(1,0)}(0,v)| \\
&\le \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{31j}(v)||\tilde{\sigma}_j^{(1,0)}(0,v)| + \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{32j}(v)||\tilde{\sigma}_j^{(1,0)}(0,v)|.
\end{aligned} \tag{A31}$$

For the first term in (A31), we have

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{31j}(v)|\left|\tilde{\sigma}_j^{(1,0)}(0,v)\right| \mathrm{d}W(v) \\
&\le 2\left[\sum_{i=1}^{n} |\hat{\varepsilon}_i - \varepsilon_i(\hat{\theta})|\right] \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \int \left|\tilde{\sigma}_j^{(1,0)}(0,v)\right| \mathrm{d}W(v) \\
&= O_P(1) O_P\left(\frac{p}{n^{1/2}}\right) = O_P\left(\frac{p}{n^{1/2}}\right),
\end{aligned} \tag{A32}$$

by (A19), Markov's inequality and the fact that $n_j E|\sigma_j^{(1,0)}(0,v)|^2 \le C$ under $\mathbb{H}_0$.

For the second term in (A31), using the second-order Taylor series expansion of $\varepsilon_i(\hat{\theta})$ in (A13), we have

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{32j}(v)|\left|\tilde{\sigma}_j^{(1,0)}(0,v)\right| \mathrm{d}W(v) \\
&\le \sqrt{n}\left\|\hat{\theta} - \theta_0\right\| n^{-1/2} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int \left\| n_j^{-1} \sum_{i=j+1}^{n} \lambda_i(\theta_0)\phi_{i-j}(v)\right\| \left|\tilde{\sigma}_j^{(1,0)}(0,v)\right| \mathrm{d}W(v) \\
&+ \left\|\sqrt{n}(\hat{\theta} - \theta_0)\right\|^2 \left[n_j^{-1} \sum_{i=1}^{n} \sup_{\theta \in \Theta_0} \left\|\lambda_i'(\theta)\right\|\right] \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \int \left|\tilde{\sigma}_j^{(1,0)}(0,v)\right| \mathrm{d}W(v) \\
&= O_P\left(\frac{1+p}{n^{1/2}}\right) + O_P\left(\frac{1+p}{n^{1/2}}\right) = o_P(p^{1/2})
\end{aligned} \tag{A33}$$

given $p/n \to \infty$. The orders of magnitude for the terms in (A33) is obtained by the following reasoning. Given (A22) and the fact that $n_j E|\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \le C$ under $\mathbb{H}_0$, we have

$$E\left[\left\|n_j^{-1}\sum_{i=j+1}^{n}\lambda_i(\theta_0)\phi_{i-j}(v)\right\|\left|\tilde{\sigma}_j^{(1,0)}(0,v)\right|\right] \le \left[E\left\|n_j^{-1}\sum_{i=j+1}^{n}\lambda_i(\theta_0)\phi_{i-j}(v)\right\|^2\right]^{1/2}\left[E\left|\tilde{\sigma}_j^{(1,0)}(0,v)\right|^2\right]^{1/2}$$
$$\le C\left[\|\xi_j(v)\| + Cn_j^{-1/2}\right]n_j^{-1/2}.$$

Consequently,

$$n^{-1/2}\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j E\int\left\|n_j^{-1}\sum_{i=j+1}^{n}\lambda_i(\theta_0)\phi_{i-j}(v)\right\|\left|\tilde{\sigma}_j^{(1,0)}(0,v)\right|\mathrm{d}W(v)$$
$$\le C\sum_{j=1}^{n-1}\int\|\xi_j(v)\|\,\mathrm{d}W(v) + Cn^{-1/2}\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right) = O(1+p/n^{1/2}),$$

given Assumptions A.5–A.7 and (A21). It follows that the first term in (A33) is $O_P(1 + p/n^{1/2})$ by Markov's inequality. The second term in (A33) is $O_P(p/n^{1/2})$ given Assumption A.4, (A15), (A11) and the fact that $n_j E|\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \le C$ under $\mathbb{H}_0$. Combining (A31)–(A33) then yields the result for $c = 3$. This completes the proof of Proposition A.2. □

Proof of Theorem A.2. The proof of Theorem A.2 is similar to that of Theorem A.1. Let $\hat{A}_{q,1}$ and $\hat{A}_{q,1}$ be defined in the same way as $\hat{A}_1$ and $\hat{A}_2$ in (A8) and (A9) respectively, with $\{\varepsilon_{q,i}\}_{i=1}^n$ replacing $\{\hat{\varepsilon}_i\}_{i=1}^n$. It suffices to show $p^{-1/2}\hat{A}_{q,1}\xrightarrow{p}0$ and $p^{-1/2}\hat{A}_{q,2}\xrightarrow{p}0$. Put $\delta_{q,i} \equiv e^{\mathbf{i}v\varepsilon_i} - e^{\mathbf{i}v\varepsilon_{q,i}}$ *and* $\phi_{q,i}(v) \equiv e^{\mathbf{i}v\varepsilon_{q,i}} - \varphi_q(v)$, where $\varphi_q(v) \equiv E(e^{\mathbf{i}v\varepsilon_{q,i}})$. Let $\tilde{\sigma}_{q,j}^{(1,0)}(0,v)$ be defined as $\tilde{\sigma}_j^{(1,0)}(0,v)$, with $\{\varepsilon_{q,i}\}_{i=1}^n$ replacing by $\{\varepsilon_i\}_{i=1}^n$. Then, we can decompose

$$\begin{aligned}&\tilde{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_{q,j}^{(1,0)}(0,v)\\ &= \mathbf{i}n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_i - \varepsilon_{q,i})\phi_{i-j}(v) + \mathbf{i}n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_{q,i}-1)\delta_{q,i-j}(v) + \mathbf{i}\left[\varphi_q(v)-\varphi(v)\right]n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_{q,i}-1)\\ &= \mathbf{i}\left[\hat{B}_{q,1j}(v) + \hat{B}_{q,2j}(v) + \hat{B}_{q,3j}(v)\right],\ \text{say.}\end{aligned} \tag{A34}$$

For the $\hat{B}_{1j,q}(v)$ term in (A34), observing that $(\varepsilon_i - \varepsilon_{q,i})\phi_{i-j}(v)$ is an m.d.s. under $\mathbb{H}_0$, we have

$$E|\hat{B}_{q,1j}(v)|^2 = n_j^{-1}E\left[(\varepsilon_i-\varepsilon_{q,i})^2|\phi_{i-j}(v)|^2\right] \le n_j^{-1}E(\varepsilon_i-\varepsilon_{q,i})^2 \le Cn_j^{-1}q^{-\kappa}$$

by Assumption A.8. It follows from Markov's inequality, (A11) and Assumptions A.7 and A.8 that

$$\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j|\hat{B}_{q,1j}(v)|^2\,\mathrm{d}W(v) = O_P\left(\frac{p}{q^\kappa}\right).$$

Next, for the $\hat{B}_{2j,q}(v)$ term in (A34), by the Cauchy–Schwarz inequality and the inequality that $|e^{\mathbf{i}z_1} - e^{\mathbf{i}z_2}| \le |z_1 - z_2|$ for any real-valued variables z_1 and z_2, we have

$$E|\hat{B}_{q,2j}(v)|^2 = n_j^{-1}E[(\varepsilon_{q,i}-1)^2|\delta_{i-j}(v)|^2] \le n_j^{-1}[E(\varepsilon_{q,i}^4)]^{1/2}\left[E(\varepsilon_{q,i}-\varepsilon_i)^4\right]^{1/2} \le Cn_j^{-1}q^{-\kappa}.$$

It follows from (A11), and Assumptions A.7 and A.8 that

$$\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j|\hat{B}_{q,2j}(v)|^2\,\mathrm{d}W(v) = O_P\left(\frac{p}{q^\kappa}\right).$$

Finally, we consider the $\hat{B}_{q,3j}(v)$ term in (A34). Because

$$|\varphi_q(v) - \varphi(v)| \le E|\delta_i(v)| \le |v|E|\varepsilon_{q,i} - \varepsilon_i|$$

and

$$E\left[n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_{q,i}-1)\right]^2 \le n_j^{-1}E(\varepsilon_{q,i}^2) \le Cn_j^{-1}$$

by the m.d.s. property of $\{\varepsilon_{q,i} - 1\}$, we have

$$E\left|\hat{B}_{q,2j}(v)\right|^2 = \left|\varphi_q(v)-\varphi(v)\right|^2 E\left[n_j^{-1}\sum_{i=j+1}^{n}(\varepsilon_{q,i}-1)\right]^2 \le Cv^2n_j^{-1}q^{-\kappa}$$

given Assumption A.8. It follows from (A11) that

$$\int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{q,3j}(v)|^2 \, dW(v) = O_P\left(\frac{p}{q^{\kappa}}\right).$$

Therefore, we obtain

$$p^{-1/2}\hat{A}_{q,1} = 16p^{-1/2} \sum_{c=1}^{3} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int \left|\hat{B}_{q,cj}(v)\right|^2 dW(v) = O_P\left(\frac{p^{1/2}}{q^{\kappa}}\right) = o_P(1)$$

given $p/q^{\kappa} \to 0$. Moreover, by (A34) and the Cauchy–Schwartz inequality, we have

$$\begin{aligned} p^{-1/2}\hat{A}_{q,2} &= 16p^{-1/2} \sum_{c=1}^{3} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \ \mathrm{Re} \int \hat{B}_{q,cj}(v) \tilde{\sigma}_{qj}^{(1,0)}(0,v)^* \, dW(v) \\ &\le p^{-1/2} \sum_{c=1}^{3} \left[\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{q,cj}(v)|^2 \, dW(v)\right]^{\frac{1}{2}} \left[\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\tilde{\sigma}_{qj}^{(1,0)}(0,v)|^2 \, dW(v)\right]^{1/2} \\ &= O_P(q^{-1/2}\kappa) O_P(p^{1/2}) = O_P\left(\frac{p^{1/2}}{q^{(1/2)}\kappa}\right) = o_P(1) \end{aligned}$$

given $p/q^{\kappa} \to 0$, where

$$\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\tilde{\sigma}_{qj}^{(1,0)}(0,v)|^2 \, dW(v) = O_P(p)$$

by Markov's inequality and the fact that $E|\tilde{\sigma}_{qj}^{(1,0)}(0,v)|^2 \le Cn_j^{-1}$. This completes the proof of Theorem A.2. □

Proof of Theorem A.3. The proof follows closely the proof of Thm A.3 in Hong and Lee (2005) with $\{\varepsilon_{q,i}, \psi_{q,i}(v)\}$ replaced by $\{\varepsilon_{q,t} - 1, \phi_{q,t}(v)\}$, given $q = p^{1+(1/(4b-2))} (ln^2 n)^{1/(2b-1)}$ and $p = cn^{\lambda}$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$. □

Proof of Theorem 2. We consider the proof of $\hat{M}_1^d(p)$ only; the proof for $\hat{M}_0^d(p)$ is similar and simpler. We shall show that $\hat{M}_1^d(p) - \hat{M}_1(p) \xrightarrow{p} 0$. The asymptotic normality of $\hat{M}_1^d(p)$ then follows immediately from Theorem 1. To show that $\hat{M}_1^d(p) - \hat{M}_1(p) \xrightarrow{p} 0$, it suffices to show that (i) $p^{-1/2}[\hat{C}_1^d(p) - \hat{C}_1(p)] = O_P(n^{-1/2})$, (ii) $p^{-1}[\hat{D}_1^d(p) - \hat{D}_1(p)] \xrightarrow{p} 0$ and (iii)

$$p^{-1/2} \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \left[|\hat{\gamma}_j^{(1,0)}(0,v)|^2 - |\hat{\sigma}_j^{(1,0)}(0,v)|^2\right] dW(v) \xrightarrow{p} 0, \tag{A35}$$

where $\hat{D}_1(p) = pD\int_0^{\infty} k^4(z)\,dz[1 + o(1)] \propto p$ as noted in the proof of Theorem A.1. For space, we focus on the proof of (A35); the proofs for (i) and (ii) here are relatively straightforward. Note that the convergence rate $O_P(n^{-1/2})$ for $p^{-1/2}[\hat{C}_1^d(p) - \hat{C}_1(p)]$ implies that replacing $\hat{C}_1^d(p)$ with $\hat{C}_1(p)$ has asymptotically negligible impact given $p/n \to 0$.

To show (A35), we decompose

$$\int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j [|\hat{\gamma}_j^{(1,0)}(0,v)|^2 - |\hat{\sigma}_j^{(1,0)}(0,v)|^2] \, dW(v) = \hat{A}_3 + 2\,\mathrm{Re}(\hat{A}_4),$$

where

$$\begin{aligned} \hat{A}_3 &= \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \left|\hat{\gamma}_j^{(1,0)}(0,v) - \hat{\sigma}_j^{(1,0)}(0,v)\right|^2 dW(v), \\ \hat{A}_4 &= \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j [\hat{\gamma}_j^{(1,0)}(0,v) - \hat{\sigma}_j^{(1,0)}(0,v)] \hat{\sigma}_j^{(1,0)}(0,v)^* \, dW(v). \end{aligned}$$

Then (A35) follows from Theorems A4 and A5, and $p \to \infty$ as $n \to \infty$. □

Theorem A.4. *Under the conditions of Theorem 2, $\hat{A}_3 = O_P(1)$, and $p^{-1/2}\hat{A}_3 \to^p 0$.*

Theorem A.5. *Under the conditions of Theorem 2, $p^{-1/2}\hat{A}_4 \to^p 0$.*

Proof of Theorem A.4. *By the definitions of* $\hat{\gamma}_j^{(1,0)}(0,v)$ *and* $\hat{\sigma}_j^{(1,0)}(0,v)$, *we have for* $j > 0$,

$$\hat{\gamma}_j^{(1,0)}(0,v) - \hat{\sigma}_j^{(1,0)}(0,v) = \mathbf{i}n_j^{-1}\sum_{i=j+1}^{n}(\hat{\varepsilon}_i - 1)\left[\hat{h}_{i-j}(v) - \hat{\phi}_{i-j}(v)\right] = -\mathbf{i}\hat{\beta}_j(v)'n_j^{-1}\sum_{i=j+1}^{n}\hat{G}_i(\hat{\varepsilon}_i - 1),$$

where $\hat{h}_{i-j}(v) \equiv \hat{\phi}_{i-j}(v) - \hat{G}_i'\hat{\beta}_j(v)$, $\hat{\beta}_j(v) \equiv (\sum_{i=1}^{n}\hat{G}_i\hat{G}_i')^{-1}\sum_{i=j+1}^{n}\hat{G}_i\hat{\phi}_{i-j}(v)$, $\hat{G}_i \equiv \frac{\partial}{\partial\theta}\ln\psi(I_{i-1}^{\dagger},\hat{\theta})$, $\hat{\phi}_{i-j}(v) \equiv e^{\mathbf{i}v\hat{\varepsilon}_{i-j}} - \hat{\varphi}(v)$ and $\hat{\varphi}(v) = n^{-1}\sum_{i=1}^{n}e^{\mathbf{i}v\hat{\varepsilon}_i}$. The proof of Theorem A.4 is analogous to the proof of Thm A.1 in Hong and Lee (2007), with $\hat{\phi}_j(v)$ and $\hat{\varphi}(v)$ here replacing $\hat{\psi}_t(v)$ and $\hat{\varphi}_j(v)$ in Hong and Lee (2007) respectively. We note that although we use the same notations, $\hat{G}_i$ and $\hat{\varepsilon}_i$ denote $\frac{\partial}{\partial\theta}\ln\psi(I_{i-1}^{\dagger},\hat{\theta})$ and $Y_i/\psi(I_{i-1}^{\dagger},\hat{\theta})$ respectively here whereas $\hat{G}_t$ and $\hat{\varepsilon}_t$ denote $\frac{\partial}{\partial\theta}g(I_{t-1}^{\dagger},\hat{\theta})$ and $Y_t - g(I_{t-1}^{\dagger},\hat{\theta})$ respectively in Hong and Lee (2007). However, the proof is similar given Assumptions A.1–A.10. The detailed proof of Theorem A.4 can be obtained from the authors on request. □

Proof of Theorem A.5. The proof is analogous to the proof of Thm A.2 in Hong and Lee (2007) given Assumptions A.1–*A*.10, with the same explanations of notations as in the proof of Theorem A.4. The detailed proof of Theorem A.5 can be obtained from the authors on request. □

Proof of Theorem 3(i). We consider the proof for $\hat{M}_1(\mathrm{p})$ only. It consists of the proofs of Theorems A.6 *and A*.7 that follow. □

Theorem A.6. *Under the conditions of Theorem* 3(*i*), $(p^{1/2}/n)[\hat{M}_1(p) - \tilde{M}_1(p)] \to^p 0$, *where* $\tilde{M}_1(p)$ *is as in Theorem A.1.*

Theorem A.7. *Let* $\tilde{M}_1(\mathrm{p})$ *be as in Theorem A.6. Under the conditions of Theorem* 3(*i*),

$$(p^{1/2}/n)\tilde{M}_1(p) \underset{p}{\to} \left[2D\int_0^{\infty}k^4(z)\,\mathrm{d}z\right]^{-1/2}\sum_{j=1}^{\infty}\int|\sigma_j^{(1,0)}(0,v)|^2\,\mathrm{d}W(v).$$

Proof of Theorem A.6. It suffices to show that

$$n^{-1}\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j\left[|\hat{\sigma}_j^{(1,0)}(0,v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0,v)|^2\right]\mathrm{d}W(v) \xrightarrow{p} 0, \tag{A36}$$

$p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)] = O_P(1)$, $p^{-1}\hat{D}_1(p) - \tilde{D}_1(p) \xrightarrow{p} 0$ and $\tilde{D}_1(p) = pD\int_0^{\infty}k^4(z)\,\mathrm{d}z[1 + o(1)]$. Since the proofs for $p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)] = O_P(1)$, $p^{-1}[\hat{D}_1(p) - \tilde{D}_1(p)] \xrightarrow{p} 0$ and $\tilde{D}_1(p) = pD\int_0^{\infty}k^4(z)\,\mathrm{d}z[1 + o(1)]$ are relatively straightforward, we focus on the proof of (A36). From (A9), the Cauchy–Schwarz inequality, and the fact that

$$n^{-1}\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j|\tilde{\sigma}_j^{(1,0)}(0,v)|^2\,\mathrm{d}W(v) = O_P(1)$$

as is implied by Theorem A.7 (the proof of Theorem A.7 does not depend on Theorem A.6), it suffices to show that $n^{-1}\hat{A}_1 \xrightarrow{p} 0$, where $\hat{A}_1$ is defined as in (A8). Given the decomposition in (A10), we shall show that

$$n^{-1}\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j|\hat{B}_{cj}(v)|^2\,\mathrm{d}W(v) \xrightarrow{p} 0$$

for $c = 1,\ldots,5$.

We first consider $c = 1$. By the Cauchy–Schwarz inequality and $|\hat{\delta}_i(v)| \le 2$, we have

$$|\hat{B}_{1j}(v)|^2 \le \left[n_j^{-1}\sum_{i=j+1}^{n}(\hat{\varepsilon}_i - \varepsilon_i)^2\right]\left[n_j^{-1}\sum_{i=j+1}^{n}|\hat{\delta}_i(v)|^2\right] \le 4n_j^{-1}\sum_{i=1}^{n}(\hat{\varepsilon}_i - \varepsilon_i)^2.$$

It follows from (A6) and (A11) that

$$n^{-1}\int\sum_{j=1}^{n-1}k^2\left(\frac{j}{p}\right)n_j|\hat{B}_{1j}(v)|^2\,\mathrm{d}W(v) \le 4\left[\sum_{i=1}^{n}(\hat{\varepsilon}_i - \varepsilon_i)^2\right]\sum_{j=1}^{n-1}a_n(j)\left[\int\mathrm{d}W(v)\right]^2 = O_P\left(\frac{p}{n}\right).$$

For $c = 2$, by the Cauchy–Schwarz inequality and the inequality that $|e^{\mathbf{i}z_1} - e^{\mathbf{i}z_2}| \le |z_1 - z_2|$ for any real-valued variables z_1 and z_2, we have

$$|\hat{B}_{2j}(v)|^2 \leq v^2 \left[n_j^{-1} \sum_{i=j+1}^{n} (\varepsilon_i - 1)^2 \right] \left[n_j^{-1} \sum_{i=j+1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)^2 \right].$$

It follows from (A6), (A11) and Assumption A.7 that

$$n^{-1} \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{2j}(v)|^2 \, dW(v) \leq \sum_{j=1}^{n-1} a_n(j) \left[\sum_{i=1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)^2 \right] \int v^2 \, dW(v) = O_P\left(\frac{p}{n}\right).$$

Next, we consider $c = 3$. By the Cauchy–Schwarz inequality and $|\phi_i(v)| \leq 2$, we have

$$|B_{3j}(v)|^2 \leq 4 n_j^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)^2.$$

It follows that

$$n^{-1} \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{3j}(v)|^2 \, dW(v) \leq 4 \left[n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)^2 \right] \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \int \, dW(v) = O_P\left(\frac{p}{n}\right).$$

For $c = 4$, given $|\hat{\varphi}(v) - \varphi(v)| \leq 2$, we have

$$\left|\hat{B}_{3j}(v)\right|^2 \leq 4 n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)^2,$$

and so

$$n^{-1} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int |\hat{B}_{4j}(v)|^2 \, dW(v) \leq 4 n^{-1} \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \left[n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \varepsilon_i)^2 \right] \int \, dW(v) = O_P\left(\frac{p}{n}\right)$$

by (A6), (A11) and Assumption A.7.
Finally, for $c = 5$, we have

$$|B_{5j}(v)|^2 \leq |\hat{\varphi}(v) - \varphi(v)|^2 n_j^{-1} \sum_{j=1}^{n-1} (\varepsilon_i - 1)^2.$$

It follows that

$$n^{-1} \int \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j |\hat{B}_{5j}(v)|^2 \, dW(v) \leq |\hat{\varphi}(v) - \varphi(v)|^2 \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \left[n^{-1} \sum_{j=1}^{n-1} (\varepsilon_i - 1)^2 \right] = O_P\left(\frac{p}{n}\right)$$

from Markov's inequality, (A11) and (A27). This completes the proof for Theorem A.6. □

PROOF OF THEOREM A.7. The proof is very similar to Hong (1999, proof of Thm 5), for the case $(m,l) = (1,0)$ and $W_1(\cdot) = \delta(\cdot)$, the Dirac delta function. □

PROOF OF THEOREM 3(II). Recall that $h_{i-j}(v) = \phi_{i-j}(v) - G_i'\beta_j(v)$, where $\beta_j = [E(G_i G_i')]^{-1} E[G_i \phi_{i-j}(v)]$. We define the following pseudo test statistic

$$\tilde{M}_1^d(p) = \frac{\left[\sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) n_j \int \left| \tilde{\gamma}_j^{(1,0)}(0,v) \right|^2 \, dW(v) - \tilde{C}_1^d(p) \right]}{\sqrt{\tilde{D}_1^d(p)}},$$

where

$$\tilde{\gamma}_j^{(1,0)}(0,v) = n_j^{-1} \sum_{i=j+1}^{n} \mathbf{i} (\varepsilon_i - 1) h_{i-j}(v),$$

and

$$\tilde{C}_1^d(p) = \sum_{j=1}^{n-1} k^2\left(\frac{j}{p}\right) \int n_j^{-1} \sum_{i=j+1}^{n} (\varepsilon_i^2 - 1)\left|h_{i-j}(v)\right|^2 \mathrm{d}W(v),$$

$$\tilde{D}_1^d(p) = \sum_{j=1}^{n-2}\sum_{l=1}^{n-2} k^2\left(\frac{j}{p}\right) k^2\left(\frac{l}{p}\right) \int\int \left|\frac{1}{n-\max(j,l)} \sum_{i=\max(j,l)+1}^{n} (\varepsilon_i^2 - 1) h_{i-j}(v) h_{i-l}(v)\right|^2 \mathrm{d}W(u)\,\mathrm{d}W(v).$$

The proof of Theorem 3(ii) consists of the proofs of Theorems A.8 and A.9, where Theorem A.8 shows that replacing the estimated residuals $\{\hat{\varepsilon}_i\}_{i=1}^n$ with the unobservable sample $\{\varepsilon_i\}_{i=1}^n$ and replacing the OLS estimators $\{\hat{\beta}_j(v)\}_{j=1}^{n-1}$ with their population counterparts $\{\beta_j(v)\}_{j=1}^{n-1}$ do not affect the asymptotic behaviour of $(p^{1/2}/n)\hat{M}_1^d(p)$ under $\mathbb{H}_A$. Theorem A.9 shows that $(p^{1/2}/n)\tilde{M}_1^d(p)$ converges to a well-defined probability limit under $\mathbb{H}_A$ from which the $\hat{M}_1^d(p)$ test gains its power. □

Theorem A.8. *Under the conditions of Theorem 3,* $(p^{1/2}/n)[\hat{M}_1^d(p) - \tilde{M}_1^d(p)] \to^p 0$.

Theorem A.9. *Under the conditions of Theorem* 3,

$$(p^{1/2}/n)\tilde{M}_1^d(p) \xrightarrow{p} \left[2D_d \int_0^\infty k^4(z)\,\mathrm{d}z\right]^{-1/2} \sum_{j=1}^{\infty} \int |\gamma_j^{(1,0)}(0,v)|^2\,\mathrm{d}W(v).$$

Proof of Theorem A.8. The proof of Theorem A.8 *is analogous to the proof of Theorem A.*3 in Hong and Lee (2007), with the same explanations for some notations as in the proof of Theorem A.3 of this article. The detailed proof is available from the authors on request. □

Proof of Theorem A.9. See Hong (1999, proof of Thm 5) for the case of $(m,l) = (1,0)$. We note that following reasoning analogous to the proof of Hong and Lee (2005, proof of Thm 1), we can obtain $\tilde{C}_1^d(\mathrm{p}) = \mathrm{O_P}(\mathrm{p})$ and $p^{-1}\tilde{D}_1^d(\mathrm{p}) \xrightarrow{\mathrm{p}} 2\mathrm{D_d} \int_0^\infty \mathrm{k}^4(\mathrm{z})\,\mathrm{dz}$, where D_d is as in Theorem 3(ii). □

Acknowledgements

The authors thank two referees, associate editor and the editor (Maurine Priestley) for insightful comments. They also thank participants of the Time Series Conference in Montreal, the Third Symposium on Econometric Theory and Applications in Hong Kong University of Science and Technology, the International Symposium on Forecasting and Risk Management at Chinese Academy of Science for comments. Yongmiao Hong thanks financial support from the Outstanding Oversea Youth Grant of the National Science Foundation of China and the Cheung Kong Scholarship by Chinese Ministry of Education and Xiamen University, China, and Yoon-Jin Lee thanks financial support from Department of Economics, Indiana University, USA.

REFERENCES

Bahadur, R. R. (1960) Stochastic comparison of tests. *Annals of Mathematical Statistics* **31**, 276–95.

Bauwen, L. and Giot, P. (2000) The logarithmic ACD model: application to the bid–ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique* **60**, 117–49.

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

Box, G. E. P. and Pierce, D. A. (1970) Distribution of the residual autocorrelations in ARIMA time series models. *Journal of the American Statistical Association* **65**, 1509–26.

Brillinger, D. (1965) An introduction to polyspectra. *Annals of Mathematical Statistics* **36**, 1351–74.

Brillinger, D. and Rosenblatt, M. (1967a) Asymptotic theory of estimates of k-th order spectra. In *Spectral Analysis of Time Series*, (ed. B. Harris). New York: John Wiley, pp. 153–88.

Brillinger, D. and Rosenblatt, M. (1967b) Computation and interpretation of k-th order spectra. In *Spectral Analysis of Time Series*, (ed. B. Harris). New York: John Wiley, pp. 189–232.

Chou, R. (2005) Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking* **37**, 561–82.

Diebold, F. X., Gunther, T. A. and Tsay, A. S. (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* **39**, 863–83.

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

Dufour, A. and Engle, R. F. (2000) *The ACD Model: Predictability of the Time between Consecutive Trades*, Discussion Paper, ISMA Centre, University of Reading.

Duchesne, P. and Pacurar, M. (2006) Evaluating financial time series models for irregularly spaced data: a spectral density approach. *Computers and Operations Research* **35**, 130–55.

Easley, D. and O'Hara, M. (1992) Time and the process of security price adjustment. *Journal of Finance* **19**, 69–90.

Easley, D., Kiefer, N. and O'Hara, M. (1997) The information content of the trading process. *Journal of Empirical Finance* **4**, 69–90.

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

Engle, R. F. (2002) New frontiers for ARCH models. *Journal of Applied Econometrics* **17**, 425–46.
Engle, R. F. and Russell, J. R. (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**, 1127–62.
Escanciano, J. (2006) Goodness–of–fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association* **101**, 531–41.
Escanciano, J. and Velasco, C. (2006) Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* **134**, 151–85.
Fernandes, M. and Grammig, J. (2005) Nonparametric specification tests for conditional duration models. *Journal of Econometrics* **127**, 35–68.
Fernandes, M. and Grammig, J. (2006) A family of autoregressive conditional duration models. *Journal of Econometrics* **130**, 1–23.
Ghysels, E., Gouriéroux, C. and Jasiak, J. (2004) Stochastic volatility durations. *Journal of Econometrics* **119**, 413–35.
Glosten, L. R. and Milgrom, P. (1985) Bid ask and transaction prices in a specialist market with heterogeneously informed agents. *Journal of Financial Economics* **14**, 71–100.
Hautsch, N. (2003) Assessing the risk of liquidity suppliers on the basis of excess demand intensities. *Journal of Financial Econometrics* **1**, 189–215.
Hautsch, N. (2006) *Testing the Conditional Mean Function of Autoregressive Conditional Duration Models*, Working Paper, University of Copenhagen.
Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–64.
Hong, Y. (1997) One–sided testing for conditional heteroskedasticity in time series models. *Journal of Time Series Analysis* **18**, 253–77.
Hong, Y. (1999) Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *Journal of the American Statistical Association* **84**, 1201–20.
Hong, Y. and Lee, T.-H. (2003) Diagnostic checking for the adequacy of nonlinear time series models. *Econometric Theory* **19**, 1065–121.
Hong, Y. and Lee, Y.-J. (2005) Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Review of Economic Studies* **72**, 499–541.
Hong, Y. and Lee, Y.-J. (2007) An improved generalized spectral test for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Econometric Theory* **72**, 499–541.
Horowitz, J. (2003) Bootstrap methods for Markov processes. *Econometrica* **71**, 1049–82.
Hujer, R., Vuletić, S. and Kokot, S. (2002) *The Markov Switching ACD Model*. Johann Wolfgang Goethe–University, Finance & Accounting Working Paper No. 90.
Jasiak, J. (1998) Persistence in intertrade durations. *Finance* **19**, 166–95.
Lancaster, T. (1990) *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
Li, W. K. and Mak, T. (1994) On the squared autocorrelations in non–linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* **15**, 627–36.
Li, W. K. and Yu, L. H. (2003) On the residual autocorrelation of the autoregressive conditional duration model. *Economics Letters* **79**, 169–75.
Liu, Y., Hong, Y. and Wang, S. (2006) *How Well Can Autoregressive Duration Models Capture the Price Duration Dynamics of Foreign Exchanges*, Working Paper, Department of Economics, Cornell University.
Meitz, M. and Teräsvirta, T. (2006) Evaluating models of autoregressive conditional duration. *Journal of Business and Economic Statistics* **24**, 104–24.
Nelson, D. B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347–70.
Pacurar, M. (2006) *Autoregressive conditional duration (ACD) models in finance: a survey of the theoretical and empirical literature*, Working Paper, Dalhousie University.
Priestley, M. (1981) *Spectral Analysis and Time Series*. London: Academic Press.
Stinchcombe, M. and White, H. (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* **14**, 295–324.
Tong, H. (1990) *Nonlinear Time Series: A Dynamic System Approach*. Oxford: Clarendon Press.
White, H. (2001) *Asymptotic Theory for Econometrician*, 2nd edn. San Diego: Academic Press.
Wooldridge, J. (1990a) A unified approach to robust, regression–based specification tests. *Econometric Theory* **6**, 17–43.
Wooldridge, J. (1990b) An encompassing approach to conditional mean tests with applications to testing nonnested hypotheses. *Journal of Econometrics* **45**, 331–50.
Wooldridge, J. (1991) On the application of robust, regression–based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* **47**, 5–46.
Zhang, M., Russell, J. and Tsay, R. (2001) A nonlinear autoregressive conditional duration model with application to financial transaction data. *Journal of Econometrics* **104**, 179–207.