JMSE **2017**, 2(1), 1–33
doi:10.3724/SP.J.1383.201001
http://www.jmse.org.cn/
http://engine.scichina.com/publisher/CSPM/journal/JMSE

Article

A General Approach to Testing Volatility Models in Time Series

Yongmiao Hong [1,2,*] and Yoon-Jin Lee [3]

[1] Department of Economics and Department of Statistical Science, Uris Hall, Cornell University, Ithaca, NY 14850, USA
[2] Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, Fujian, China
[3] Department of Economics, Kansas State University, Waters Hall, Manhattan, KS 66506, USA; yoonjin@ksu.edu
* Correspondence: yh20@cornell.edu; Tel.: +1-607-255-5130

Received: 12 August 2016; Accepted: 27 March 2017; Published: 17 April 2017

Abstract: Volatility models have been playing important roles in economics and finance. Using a generalized spectral second order derivative approach, we propose a new class of generally applicable omnibus tests for the adequacy of linear and nonlinear volatility models. Our tests have a convenient asymptotic null N(0,1) distribution, and can detect a wide range of misspecifications for volatility dynamics, including both neglected linear and nonlinear volatility dynamics. Distinct from the existing diagnostic tests for volatility models, our tests are robust to time-varying higher order moments of unknown form (e.g., time-varying skewness and kurtosis). They check a large number of lags and are therefore expected to be powerful against neglected volatility dynamics that occurs at higher order lags or display long memory properties. Despite using a large number of lags, our tests do not suffer much from the loss of a large number of degrees of freedom, because our approach naturally discounts higher order lags, which is consistent with the stylized fact that economic or financial markets are affected more by the recent past events than by the remote past events. No specific estimation method is required, and parameter estimation uncertainty has no impact on the convenient limit N(0,1) distribution of the test statistics. Moreover, there is no need to formulate an alternative volatility model, and only estimated standardized residuals are needed to implement our tests. We do not have to calculate tedious and model-specific score functions or derivatives of volatility models with respect to estimated parameters, which are required in some existing popular diagnostic tests for volatility models. We examine the finite sample performance of the proposed tests. It is documented that the new tests are rather powerful in detecting neglected nonlinear volatility dynamics which the existing tests can easily miss. They are useful diagnostic tools for practitioners when modelling volatility dynamics.

Keywords: GARCH models; Nonlinear volatility dynamics; Specification testing

1. Introduction

Volatility is one of the most important instruments in economics and finance. As volatility represents the degree of unexpected price movement over time, volatility modeling and forecasting is important in investment, security valuation, risk management, and monetary policymaking. As a measure of uncertainty, volatility is a key input into many investment decisions and portfolio choices. It thus plays a central role in the pricing of derivative securities, where the uncertainty associated with the future price of the underlying asset is the most important determinant of derivative prices. In addition, scholars can investigate how information transmits across financial markets by examining volatility spillovers, while understanding the volatility transmission mechanism between asset prices and GDP growth is important for policymakers to reduce output volatility. Policymakers often rely on market estimates of volatility as a barometer of the vulnerability of financial markets and the overall economy.

Since the seminal works by Bollerslev (1986) and Engle (1982), traditional time series tools such as autoregressive moving average (ARMA) models for the conditional mean have been extended to essentially analogous models for conditional variance that model persistence in volatility shocks. The autoregressive conditional heteroskedasticity (ARCH) family of models are now commonly used to capture the volatility dynamics of financial time series. This class includes the ARCH and GARCH models of Bollerslev (1986) and Engle (1982) as well as their various nonlinear generalizations such as Glosten et al.'s (1993) threshold GARCH model, Higgins and Bera's (1992) nonlinear GARCH models, Nelson's (1991) EGARCH model, Sentana's (1995) quadratic GARCH model, and Zakoian's (1994) threshold ARCH model. For a survey of ARCH models, see Bera and Higgins (1993) and Bollerslev et al. (1992 & 1994).

When modeling the conditional mean, the estimated model is regularly subjected to a battery of misspecification tests to check its adequacy (e.g., check serial correlation, functional form misspecification, and parameter instability). For volatility models, testing the adequacy of the estimated model has been much less common in practice. In many cases, the ease of estimation is the primary factor affecting the choice of model, and robustness checks that analyze the sensitivity of economic results to the model specification are often omitted. However, consistent parameter estimation, optimal volatility forecasting, valid hypothesis testing, and economic interpretations all require the correct specification of volatility models.

Three categories of specification tests for GARCH models exist in the literature: Box–Pierce–Lung-type portmanteau tests, Lagrange multiplier (LM) tests, and residual-based diagnostics. Box–Pierce–Ljung-type tests for the squared standardized residuals of a GARCH model have been used to test the adequacy of the GARCH model. As these test statistics are readily computable from the standardized residuals of a GARCH model, they have been widely used in practice (e.g., Tsay, 2002, pp.115–118), with the asymptotic χ^2, distribution often used. However, in an important study, Li and Mak (1994) showed that Box–Pierce-type tests for volatility models are generally not asymptotically χ^2, because the limit distribution depends on parameter estimation uncertainty. In other words, substituting the estimated standardized residuals for the unobserved standardized innovations will change the asymptotic distribution of the test statistic. Given the foregoing, Li and Mak (1994) modified Box–Pierce–Ljung-type tests and derived the asymptotic distribution of their modified tests for GARCH models that properly takes into account the impact of parameter estimation uncertainty.

Bollerslev (1986), Bollerslev and Engle (1988), Engle and Ng (1993), and Lundbergh and Teräsvirta (2002) all proposed LM tests for GARCH models. Bollerslev (1986) suggested an LM test for testing a GARCH model against a higher order GARCH model. Lundbergh and Teräsvirta's (2002) LM test is a test of the standardized innovations being *i.i.d.* against a parametric ARCH alternative. Under normality, their test of no remaining ARCH is asymptotically equivalent to Li and Mak's (1994) test. The LM test has an advantage over portmanteau tests because of its efficiency when the alternative hypothesis is correct. Tse (2002) proposed residual-based diagnostic tests for GARCH models. These tests resort to a convenient auxiliary autoregression based on the squared standardized residuals or the cross products of the standardized estimated residuals as dependent variables, while the lagged squared standardized residuals or lagged cross products of the standardized residuals are the independent variables. Thus, as pointed out in Tse (2002), the form of regression depends on the particular type of model inadequacy the researcher wishes to investigate, which dictates the power of the tests.

From a theoretical point of view, Box–Pierce–Ljung-type tests and residual-based tests for GARCH models can detect many misspecifications in volatility dynamics. However, they can only capture linear ARCH alternatives and may miss important nonlinear volatility dynamics, especially those with zero autocorrelation in the squared standardized residuals. In particular, they may overlook certain volatility dynamics such as asymmetric behaviors in volatility.[1] We note that LM tests can be designed to detect specific nonlinear volatility features depending on the formulation of the alternative model (see Engle and Ng, 1993). In addition, existing tests for GARCH models usually check a fixed finite number of lag orders. However, recent empirical studies (e.g., Baillie et al., 1996) find that high-frequency financial time series may display a long memory in volatility clustering, where volatility depends on a very long past history. Indeed, it is an important feature of a non-Markovian process that volatility may depend on the entire past history rather than only its first few lags. Thus, it is important to check not only the functional forms of volatility dynamics but also the lag structure since neglected volatility may occur at higher order lags.

A volatility model with *i.i.d.* standardized innovations is called a strong form volatility model in the literature (Drost and Nijman, 1993). This is often assumed in practice, with either normal or t-distributions. However, a volatility model may be correctly specified when the standardized innovation displays higher order dependence possibly of unknown form. Drost and Nijman (1993) showed that even if the standardized innovation is *i.i.d.* at a certain sample frequency, when aggregated to a lower sample frequency it will become serially dependent, although it may be a martingale difference sequence (*m.d.s.*).[2] That is, the strong form volatility model is not closed under temporal aggregation and only aggregate to a weak form volatility model. A volatility model where the innovation is not *i.i.d*. is called a semi-strong or weak form volatility model, where the innovation process is an *m.d.s.* or serially uncorrelated (white noise), respectively.

1 Mikosch and Starica (2000) showed that in the case of GARCH(1,1) modeling, the sample autocorrelation function can be an extremely problematic statistical instrument that has to be used with caution when making statistical statements.

2 Ignoring serial dependence in $\{z_t\}$ by assuming *i.i.d.* will not lead to inconsistent parameter estimation for the mean and variance parameters, although it would complicate attempts to construct asymptotically efficient semiparametric estimators of the variance parameters (see Gallant and Tauchen, 1989). Lee and Hansen (1994) explicitly considered the quasi-maximum likelihood estimation (QMLE) method with *m.d.s.* innovations. Bollerslev and Wooldridge (1992) and Weiss (1986) showed that the QMLE based on strong GARCH and conditional normal distributions is consistent if the conditional variance of the semi-strong GARCH process is correctly specified.

Important studies (e.g., Gallant et al., 1991; Hansen, 1994; Harvey and Siddique, 1999 & 2000; Jondeau and Rockinger, 2003) have documented that the conditional skewness and kurtosis of asset returns are time-varying. Evidence in favor of heavy tails has also been uncovered by Gallant et al. (1991), Geweke (1994), and Jacquier et al. (2004).[3] Indeed, financial time series have been characterized with asymmetric and heavy tailed non-Gaussian distributions of unknown form and these non-Gaussian features are expected to depend on market conditions. The *i.i.d.* assumption for standardized innovations does not lead to an inconsistent estimation of a volatility model, but it affects the efficiency of parameter estimators for a semi-strong or weak form volatility model. Moreover, when constructing diagnostic tests for volatility models, it is important to take into account the impact of the time-varying higher order conditional moments of unknown form because ignoring these will lead to incorrect sizes (i.e., Type I errors). All existing tests for volatility models assume *i.i.d.* (possibly non-Gaussian) innovations and are therefore not robust to the time-varying higher order dependencies that may generate heavy tails and jumps, for example. Therefore, the extension of existing tests to allow for time-varying higher order conditional moments is crucial.

In this study, we propose a new class of diagnostic tests for the adequacy of a volatility model by using a second-order generalized spectral derivative approach. The generalized spectrum originally proposed by Hong (1999) is a frequency domain analytic tool for nonlinear time series. Intuitively, it is a spectral analysis of time series transformed via the characteristic function. Because of the use of the characteristic function, the generalized spectrum can detect all the pairwise serial dependence, including nonlinearly dependent processes that have zero autocorrelation. Unfortunately, the generalized spectrum itself is unsuitable for testing volatility models because it can capture the serial dependence not only in the conditional variance but also in the higher order conditional moments. However, we can use the second-order derivative of the generalized spectrum, which focuses only on the serial dependence in the conditional variance, making it suitable for testing volatility models.

Specifically, the second-order generalized spectral derivative is a flat function of frequency when a volatility model is correctly specified, and our test is constructed based on this fact. Because the second-order generalized spectral derivative of a correctly specified volatility model is always flat regardless of the serial dependence in higher order moments, our test can be made robust to any time-varying higher order dependence of unknown form. Our approach can check a variety of linear and nonlinear functional form misspecifications in volatility dynamics. Moreover, our frequency domain approach can check a growing number of lags as sample size increases without suffering from the "curse of dimensionality." Thus, our test is expected to be powerful against long memory volatility processes, such as the fractionally integrated GARCH model (Baillie et al., 1996). While the loss of a large number of degrees of freedom would usually result in a loss of power, this is not the case for our tests thanks to the downward weighting scheme for the higher order lags, which is consistent with the stylized fact that economic and financial markets are usually more influenced by recent events than by remote past events. The older the information, the less its impacts on current volatility.

When constructing our tests, we do not require the formulation or any prior knowledge of the

3 Jacquier et al. (2004) applied their extended SV model to equity index and exchange rate time series and found that all but one of the financial time series show strong evidence of fat-tailed errors, although weaker for the weekly series. All equity indices, weekly and daily, display a leverage effect.

alternative volatility model. Moreover, our tests are robust to parameter estimation uncertainty; in other words, the use of *estimated* standardized residuals in place of true unobservable innovations does not affect the limit distribution of the test statistics. Any $\sqrt{T}$ -consistent parameter estimator suffices. Further, no specific method is required for the estimation and only estimated standardized residuals are needed to implement our tests. Moreover, we do not need to compute the tedious case-by-case score functions or derivatives of volatility models, which are required in some popular tests for volatility models such as Li and Mak (1994) and Berkes et al. (2003). All these desirable features of our tests yield a convenient procedure in practice.

The remainder of the paper is organized as follows. Section 2 introduces the volatility models and hypotheses of interest. Section 3 introduces the generalized spectral analysis and shows how the second-order derivatives of the generalized spectral density approach can be used to test the volatility models. We derive the asymptotic normal distribution of the proposed test statistics in Section 4 and establish their asymptotic power property in Section 5. Section 6 discusses the choice of a data-driven lag order. Section 7 examines the finite sample performance of the tests by using Monte Carlo experiments. Section 8 concludes. All the mathematical proofs are presented in Appendix A. 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 $T \to \infty$. The GAUSS code to implement our tests is available from the authors upon request.

2. Hypotheses and Literature Review

Consider a stochastic time series process $\{Y_t\}$ with the following general structure:

$$\begin{cases} Y_t = \mu_t + \varepsilon_t, \\ \varepsilon_t = h_t^{1/2} z_t, \\ E(z_t \mid I_{t-1}) = 0 \;\; a.s., \\ \mathrm{Var}(z_t \mid I_{t-1}) = 1 \;\; a.s., \end{cases} \tag{2.1}$$

where $\{z_t\}$ is a sequence of unobservable *m.d.s.* innovations. By construction, $\mu_t = E(Y_t \mid I_{t-1})$ is the conditional mean of Y_t given the information set I_{t-1} available at time $t-1$ (a σ -field generated by all past observations up to time $t-1$), and $h_t = \mathrm{var}(Y_t \mid I_{t-1})$ is the conditional variance of Y_t given I_{t-1}. Both μ_t and h_t are measurable functions with respect to information set I_{t-1}. Note that I_{t-1} may include not only lagged dependent variables but also exogenous variables and may date back to the infinite remote past. An important feature of most economic and financial time series is that μ_t and h_t may depend on the entire past history of Y_t rather than only a few lags of Y_t, as is the case for ARMA and/or GARCH processes. We note that the conditions of $E(z_t \mid I_{t-1}) = 0$ and $\mathrm{var}(z_t \mid I_{t-1}) = 1$ ensure that μ_t completely captures the conditional mean dynamics of Y_t and that h_t completely captures the conditional variance and the conditional correlations of Y_t. All the serial dependence beyond the first two conditional moments is contained in $\{z_t\}$.

In many economic and financial applications, interest has been in modeling conditional variance h_t, which characterizes the volatility clustering dynamics. Important examples of volatility models include

Bollerslev's (1986) GARCH model, Ding et al.'s (1993) asymmetric power ARCH model, Glosten et al.'s (1993) asymmetric model, Higgins and Bera's (1992) nonlinear ARCH model, Nelson's (1991) exponential GARCH model, and Zakoian's (1994) threshold ARCH model.

Our interest is in checking whether a parametric volatility model $h_t(\theta) \equiv h(I_{t-1}, \theta)$ is correctly specified for $\mathrm{var}(Y_t \mid I_{t-1})$, when $.\theta \in \Theta \subset R^p.$ is a finite dimensional parameter and Θ is the parameter space. The hypotheses of interest are

$$\mathrm{H}_0 : \Pr[h_t(\theta_0) = \mathrm{Var}(Y_t \mid I_{t-1})] = 1 \ \text{ forsome} \theta_0 \in \Theta$$

versus

$$\mathrm{H}_A : \sup_{\theta \in \Theta} \Pr[h_t(\theta) \neq \mathrm{Var}(Y_t \mid I_{t-1})] > 0.$$

Because many statistical inferences for economic and financial data are based on the model $h_t(\theta)$, a test of H_0 is important from both theoretical and practical points of view. In economics and finance, interest in H_0 is often based on the assumption that the conditional mean $\mu_t(\theta)$ has been correctly specified. Thus, strictly speaking, when H_0 is rejected, it may be due to the misspecification of $h_t(\theta)$ and/or $\mu_t(\theta)$. However, for high-frequency economic and financial time series, it is often believed that there exists mild or little serial dependence in the conditional mean. Therefore, the primary focus has been on modeling the volatility of $\{Y_t\}$.

Box–Pierce–Ljung-type tests for the squared standardized residuals can be used to test the adequacy of a GARCH model $h_t(\theta)$, i.e.,

$$BP_2(p) = T(T+2)_{j=1}^{p}(T-j)^{-1}\hat{\rho}_2^2(j),\ p \in \mathrm{N}, \tag{2.2}$$

where $\hat{\rho}_2(j)$ is the sample autocorrelation function of the squared standardized residuals $\{z_t^2(\hat{\theta})\}$,

$$z_t(\hat{\theta}) = h_t(\hat{\theta})^{-1/2}\varepsilon_t(\hat{\theta}),$$

and $\hat{\theta}$ is an estimator of θ_0. As this test statistic is readily computable from the standardized residuals $z_t(\hat{\theta})$, it has been widely used with an asymptotic χ_p^2 distribution in practice (e.g., Hafner, 1998, p.112; Tsay, 2002, pp.115–118). However, Li and Mak (1994) showed that the Box–Pierce–Ljung-type test $BP_2(p)$ is generally not asymptotically χ_p^2, because the limit distribution of $BP_2(p)$ depends on parameter estimation uncertainty.[4] In other words, substituting the estimated residuals for the unobserved residuals will change the asymptotic distribution of the test statistic. Thus, it is necessary to modify the test statistics to take into account the impact of parameter estimation uncertainty.

To develop portmanteau tests based on the autocorrelations of the squared residuals in a large class of heteroskedastic time series models, Li and Mak (1994) proposed a modified Box–Pierce-type test, which explicitly takes into account the impact of parameter estimation uncertainty on the asymptotic distribution of the test:

$$Q(p) = T\hat{\rho}_2'\hat{V}^{-1}\hat{\rho}_2, \tag{2.3}$$

where $\hat{\rho}_2 = [\hat{\rho}_2(1), \cdots, \hat{\rho}_2(p)]', \hat{\rho}_2(j)$ is the sample autocorrelation in $\{z_t^2(\hat{\theta})\}$, and

$$\hat{V} = I_p - \hat{C}_0^{-2}\hat{X}\hat{G}\hat{X}',$$

4 Although the asymptotic distribution of the Box–Pierce statistics has not been firmly established, there have been arguments that the χ^2 distribution may be used as an approximation.

for some suitable matrices $\hat{X}$ and $\hat{G}$ involving the derivatives of $h(I_{t-1},\theta)$ and $\mu(I_{t-1},\theta)$ with respect to parameter θ. The $Q(p)$ test will be asymptotically χ_p^2 under H_0. Li and Mak's (1994) test is in principle a general misspecification test, but it also has an LM interpretation (see Lundbergh and Teräsvirta, 2002, pp.419–422 for more discussion). Li and Mak (1994) used the QMLE to avoid the adjustment of the degrees of freedom of the asymptotic χ_p^2 distribution according to the number of estimated parameters in $\mu_t(\theta)$.[5] However, Horváth and Kokoszka (2001) pointed out that any $\sqrt{T}$ -consistent parameter estimator will suffice. Further, Li and Mak's (1994) test statistic is model-dependent because the asymptotic variance estimator $\hat{V}$ involves the derivatives of the volatility model $h_t(\theta)$ with respect to the estimated parameter $\hat{\theta}$. They also assume conditional normality and a finite fourth moment of the observations.

$Q(p)$ is applicable to various GARCH-family models. When the null model is an $\mathrm{ARCH}(m)$, Li and Mak (1994) proposed a simpler test statistic:

$$Q(m,p)=T\sum_{j=m+1}^{p}\hat{\rho}_2^2(j),$$

and showed that it is asymptotically χ^2 distributed with $p-m$ degrees of freedom.

Horváth and Kokoszka (2001) developed asymptotic theory for a linear statistic of the sample autocorrelations of squared residuals from an $\mathrm{ARCH}(m)$ model:

$$Q_T=T\sum_{1\le j\le p(T)}\lambda_j\hat{\rho}_2^2(j),$$

where $1\le p(T)\le T$ and $p(T)\to\infty$, and $\lambda_1,\lambda_2,\cdots$ are nonnegative weights satisfying $\sum_{1\le j\le p(T)}\lambda_j<\infty$, by adding a correction term to the lag j squared residual correlation, i.e.,

$$\hat{\rho}_2(j)=\frac{1}{\hat{c}_T}\sum_{\max(j,m)<i\le T}(\hat{\varepsilon}_i^2-1)(\hat{\varepsilon}_{i-j}^2-1)-T(\theta-\hat{\theta})'C_{j,T},\ j=1,\cdots,T,$$

where $\hat{c}_T$ and $C_{j,T}$ are as given in Horváth and Kokoszka (2001). They showed that by letting $p(T)\to\infty$ and judiciously choosing the weights λ_j, limits that can be expressed as simple functionals of a Brownian motion and having well-known distributions are obtained. However, their statistic Q_T involves unknown parameter θ, and thus cannot be used in practice.

Tse (2002) proposed residual-based diagnostic tests for GARCH models. These tests can be conveniently carried out by using an artificial autoregression procedure with the squared standardized residuals or the cross products of the standardized residuals as dependent variables, and the lagged squared standardized residuals or lagged cross products of the standardized residuals as the independent variables. Thus, to a certain extent, the form of the regression depends on the particular type of model inadequacy in which the researcher is interested.

From a theoretical point of view, the above correlation-based tests can detect many misspecifications of

5 This is not the case for residuals of linear sequences, say, ARMA sequences. Box and Pierce (1970) showed that if the parameters of an $\mathrm{AR}(m)$ model are estimated by least squares, the vector of the first p residual autocorrelations has a degenerate normal distribution supported on a $(p-m)$ -dimensional hyperplane in R^p, and they concluded that the statistic $Q=T\sum_{j=1}^{p}\hat{\rho}_2^2(j)$ has a χ^2 distribution with $p-m$ degrees of freedom (see also Horváth and Kokoszka, 2001).

practical importance. However, they can only capture linear volatility alternatives and may miss important nonlinear volatility alternatives, especially those with zero autocorrelation in $z_t^2(\theta)$. For example, they can easily miss asymmetric dynamic patterns in volatility dynamics such as those that may arise because of "leverage effects" or asymmetric business cycles (e.g., Hamilton and Lin, 1996).

On the contrary, most existing tests for volatility models specify a priori an upper limit on the number of lags. From a theoretical perspective, such tests can easily miss volatility misspecification that occurs at higher order lags. In the ARCH literature, numerous estimates of GARCH models for stock markets, commodities, foreign exchange, and other asset price series have been found to be consistent with an IGARCH specification. In addition, empirical studies (e.g., Baillie et al., 1996) find that high-frequency financial time series display a long memory in volatility clustering, where h_t depends on a very long history of Y_t. Some market micro structure theory (e.g., Easley and O'Hara, 1992) suggests that high-frequency asset price volatility is a non-Markovian process. An important feature of a non-Markovian time series process is that h_t depends on the entire past history of Y_t rather than only its first few lags. Thus, it may suffer from substantial power loss by using a fixed finite lag order. In practice, one can employ a large lag order when a large sample size is available. However, the use of a large lag order usually induces the loss of a large number of degrees of freedom, causing low power against many alternatives of practical importance. In particular, today's volatility is usually influenced more by recent market events than remote market events. As a consequence, the strength of dependence in $z_t^2(\theta)$ on the past history will decay to zero as the lag order increases as is the case even for long memory volatility models.

Another drawback of existing tests for volatility models is that they all assume that standardized innovations $\{z_t\}$ are *i.i.d.* They are therefore not robust to time-varying higher order conditional moments, which are more the rule than the exception for most high-frequency financial time series. Moreover, it seems to be not obvious to robustify these tests when the serial dependence in the higher order conditional moments (e.g., skewness and kurtosis) has an unknown form. Based on the foregoing, we propose a new generally applicable test for H_0 that avoids the aforementioned undesirable features of the existing tests for volatility models.

3. Generalized Spectral Derivative Approach

3.1. Generalized Spectral Approach

We propose a new test for H_0 by using the generalized spectral approach of Hong (1999). The generalized spectrum is a spectral analysis based on the characteristic function. It is a basic frequency domain analytic tool for nonlinear time series, just as the power spectrum is a basic analytic tool for linear time series (e.g., Priestley, 1981). Both time domain and frequency domain analytic tools contain the same amount of information on the serial dependence of a time series. In some applications, however, frequency domain analysis is more enlightening and suitable. For example, as discussed below, the generalized spectrum can reveal useful information on cyclical dynamics in volatility clustering due to linear or nonlinear dependencies.

Define the standardized error of a volatility model $h_t(\theta)$,

$$z_t(\theta) = h_t^{-1/2}(\theta)\varepsilon_t(\theta), \tag{3.1}$$

where $\varepsilon_t(\theta) = Y_t - \mu_t(\theta)$. Then, H_0 is equivalent to the hypothesis that

$$\mathrm{var}\left[z_t(\theta_0) \mid I_{t-1}\right] = 1 \quad a.s. \text{for some } \theta_0 \in \Theta, \tag{3.2}$$

where I_{t-1} is the observable information set available at time $t-1$. This implies

$$\mathrm{var}\left[z_t(\theta_0) \mid I_{t-1}^z\right] = 1 \quad a.s. \text{for some } \theta_0 \in \Theta, \tag{3.3}$$

where $I_{t-1}^z \equiv \sigma(z_\tau(\theta_0), \tau \le t-1)$. Thus, to test H_0 we can check if (3.3) holds. This is essentially a standardized residual-based approach. It is convenient because there is no need to compute tedious derivatives as in Li and Mak's (1994) test. Here, we encounter the curse of dimensionality problem because I_{t-1}^z has an infinite dimension. Fortunately, the generalized spectral approach provides a sensible way in which to tackle this difficulty.

Most existing tests for volatility models are based on the sample autocorrelations in $\{z_t^2(\theta)\}$. This can only detect misspecifications in volatility model $h_t(\theta)$ that render nonzero autocorrelations in $\{z_t^2(\theta)\}$. Because autocorrelation is a measure of linear association, it may have low power compared with nonlinear volatility alternatives. Indeed, nonlinear volatility dynamics are common in practice. For example, volatility reacts differently to large price increases and large price drops, thereby producing a leverage effect in which negative (positive) shocks to the mean are associated with increases (decreases) in volatility. Further, stock price volatility tends to be higher during a recession and lower during an expansion (e.g., Hamilton and Lin, 1996). This can also generate asymmetric volatility clustering because of asymmetric business cycles (i.e., expansions last longer than recessions). Therefore, it is highly desirable to develop a test that can check a volatility model against a variety of linear and nonlinear alternatives.

To detect both linear and nonlinear departures, a sensible approach to testing H_0 is to consider a test based on a smoothed nonparametric regression estimator for $\mathrm{var}[z_t(\theta) \mid z_{t-j}(\theta)]$ for each lag order j and check whether this estimator is significantly different from one. Such a test can detect many neglected or misspecified nonlinear volatility dynamics. However, this time domain nonparametric approach does not deal with the problem associated with lag orders. Obviously, the use of a finite number of lags will render a test unable to detect the misspecification of volatility models that occurs at higher order lags.

In this study, we propose a test for H_0 by using a second-order generalized spectral derivative approach, which can investigate the linear and nonlinear serial dependence in the conditional variance. Our frequency domain approach has a number of advantages. First, our test is of nonparametric nature, and therefore is able to detect both linear and nonlinear volatility alternatives. Second, our frequency domain approach naturally incorporates information from many lags. In other words, we can test a large number of lags without suffering from the curse of dimensionality. This is particularly appealing for detecting long memory volatility alternatives. Moreover, our nonparametric approach naturally discounts higher order lags, thus alleviating the loss of a large number of degrees of freedom because of the use of many lags. As a consequence, our test is expected to be powerful compared with alternatives where the dependence in volatility decays to zero as $j \to \infty$. This is consistent with the stylized fact that today's financial markets are usually more influenced by recent events than remote past events. Another appealing feature of our approach is that no $\sqrt{T}$ -consistent parameter, $\hat{\theta}$ say, affects the limit distribution of our test statistic, which is $N(0,1)$ under H_0 One can proceed as if the true parameter θ_0 were known and were equal to for any $\sqrt{T}$ -consistent estimator. This approach thus provides a convenient procedure in practice.

For notational economy, we put $z_t \equiv z_t(\theta_0)$, where $\theta_0 = p\lim\hat{\theta}$. Suppose $\{z_t\}$ is a strictly stationary process with a marginal characteristic function $\phi(u) \equiv E(e^{\mathbf{i}uz_t})$ and pairwise joint characteristic function $\phi_j(u,v) \equiv E(e^{\mathbf{i}uz_t + \mathbf{i}vz_{t-|j|}})$, where $i \equiv \sqrt{-1}$, $(u,v) \in \mathrm{R}\times\mathrm{R}$, and $j \in \{0, \pm 1, \cdots\}$. The generalized spectrum of Hong (1999) is defined as

$$f(\omega,u,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j(u,v)e^{-\mathbf{i}j\omega},\ \ \omega \in -\pi,\pi], \tag{3.4}$$

where ω is a frequency and $\sigma_j(u,v)$ is the covariance function of the transformed series:

$$\sigma_j(u,v) \equiv \mathrm{cov}(e^{\mathbf{i}uz_t}, e^{\mathbf{i}vz_{t-|j|}}),\quad j \in \{0,\pm 1,\cdots\}. \tag{3.5}$$

Note that $f(\omega,u,v)$ is a complex-valued function. Compared with conventional power spectral density (e.g., Priestley, 1981) and higher order spectra (e.g., Brillinger, 1981), an appealing feature of $f(\omega,u,v)$ is the that no moment condition on $\{z_t\}$ is required. The function $f(\omega,u,v)$ can capture any type of pairwise serial dependence in $\{z_t\}$, i.e., dependence between z_t and z_{t-j} for any nonzero lag j, including that with zero autocorrelation. This may be called the generalized spectrum of $\{z_t\}$ because when $E(z_t^2) < \infty$, it can be differentiated to obtain the conventional power spectral density as a special case:

$$-\frac{\partial^2}{\partial u \partial v} f(\omega,u,v)\Big|_{(u,v)=(0,0)} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\mathrm{cov}(z_t, z_{t-|j|})e^{-\mathbf{i}j\omega},\ \ \omega \in -\pi,\pi], u,v \in \mathrm{R}, \tag{3.6}$$

where $\mathrm{cov}(z_t, z_{t-|j|})$ is the autocovariance of $\{z_t\}$ at lag $|j|$.

However, the generalized spectrum $f(\omega,u,v)$ itself is unsuitable for testing R_0 because it can capture the serial dependence not only in the variance but also in the higher order conditional moments of z_t. An example is when $\{z_t\}$ follows a generalized asymmetric Student t distribution with time-varying skewness and kurtosis (e.g., Hansen, 1994). In this case, $\{z_t\}$ is an *m.d.s.* process but is not *i.i.d.* The generalized spectrum $f(\omega,u,v)$ can capture this process, although $\{z_t\}$ is an *m.d.s.* with conditionally homoskedastic errors (i.e., $E(z_t \mid I_{t-1}^z) = 0$ *a.s.* and $\mathrm{var}(z_t \mid I_{t-1}^z) = 1$ *a.s.*).

However, just as the characteristic function can be differentiated to generate various moments of $\{z_t\}$, $f(\omega,u,v)$ can be differentiated to capture the serial dependence in various conditional moments. To check the serial dependence in the volatility of z_t, we can thus differentiate $f(\omega,u,v)$ and use the following second-order generalized spectral derivative:

$$f^{(0,2,0)}(\omega,0,v) \equiv \frac{\partial^2}{\partial u^2} f(\omega,u,v)\Big|_{u=0} = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j^{(2,0)}(0,v)e^{-\mathbf{i}j\omega}, \omega \in -\pi,\pi], \tag{3.7}$$

where

$$\sigma_j^{(2,0)}(0,v) \equiv \frac{\partial^2}{\partial u^2}\sigma_j(u,v)\Big|_{u=0} = -\mathrm{cov}(z_t^2, e^{\mathbf{i}vz_{t-|j|}}). \tag{3.8}$$

The measure $\sigma_j^{(2,0)}(0,v)$ exclusively focuses on the conditional variance dynamics of $\{z_t\}$. It checks whether the autoregression function $\mathrm{var}(z_t \mid z_{t-j})$ at lag j is constant. Under appropriate conditions,

$\sigma_j^{(2,0)}(0,v)=0$ for all $v \in \mathrm{R}$ if and only if $\mathrm{var}(z_t \mid z_{t-j})$ is a constant.[6] Unlike a smoothed nonparametric estimator for $\mathrm{var}(z_t \mid z_{t-j})$, $\sigma_j^{(2,0)}(0,v)$ does not involve any smoothed parameter and does not suffer from the curse of dimensionality. Moreover, the function $f^{(0,2,0)}(\omega,0,v)$ incorporates information on all lags, which is difficult to handle by using a time domain approach.

The hypothesis of $\mathrm{var}(z_t \mid I_{t-1}^z)=1$ *a.s.* is not equivalent to the hypothesis of $\mathrm{var}(z_t \mid z_{t-j})=1$ for all $j>0$. The former implies the latter but not vice versa. There exists a gap between them. This is the price we have to pay to deal with the difficulty of the curse of dimensionality. Nevertheless, the examples for which $\mathrm{var}(z_t \mid z_{t-j})=1$ for all $j>0$ but $\mathrm{var}(z_t \mid I_{t-j}^z)\neq 1$ may be rare in practice and are thus pathological.[7]

There is another payoff of using $f^{(0,2,0)}(\omega,0,v)$. Define the supremum generalized spectral derivative modulus as

$$m(\omega)\equiv \sup_{v\in \mathrm{R}}\left| f^{(0,2,0)}(\omega,0,v)\right|,\ \ \omega\in[-\pi,\pi]. \tag{3.9}$$

This can be viewed as the maximum dependence in the variance of $\{z_t\}$ at the frequency ω. It can capture the cyclical dynamics caused by either the linear or the nonlinear serial dependence in the volatility of Y_t. For example, volatility tends to be higher during a recession period than an expansion period (e.g., Hamilton and Lin, 1996). Such a cyclical pattern in volatility clustering can be easily captured by $m(\omega)$. By contrast, it may not be easily captured by the conventional power spectrum of $\{z_t^2\}$ because of the asymmetric nature of business cycles (i.e., expansion periods last longer than recession periods).

Under H_0 the second-order generalized spectral derivative $f^{(0,2,0)}(\omega,0,v)$ becomes

$$f_0^{(0,2,0)}(\omega,0,v)\equiv \frac{1}{2\pi}\sigma_0^{(2,0)}(0,v)=-\frac{1}{2\pi}\mathrm{cov}(z_t^2,e^{\mathrm{i}vz_t}),\ \ \omega\in[-\pi,\pi]. \tag{3.10}$$

This is a "flat" second-order generalized spectral derivative in the sense that $f_0^{(0,2,0)}(\omega,0,v)$ does not depend on the frequency ω ; it only depends on v. One can test H_0 by comparing two consistent estimators, namely one for $f^{(0,2,0)}(\omega,0,v)$ and another for $f_0^{(0,2,0)}(\omega,0,v)$. Any significant deviation between these estimators will indicate the rejection of H_0 The generalized spectral derivative $f^{(0,2,0)}(\omega,0,v)$ is always flat under H_0, regardless of the existence of time-varying higher order conditional moments. Therefore, we can construct a test for the adequacy of a volatility model robust to the time-varying higher order conditional moments of unknown form. Below, we use a kernel method to develop a new class of tests for H_0

3.2. Test Statistics

Suppose we have a random sample of size T and $\hat{\theta}$ is any $\sqrt{T}$-consistent estimator for θ_0. An example of $\hat{\theta}$ is the quasi-maximum likelihood estimator (e.g., Bollerslev and Wooldridge, 1992, Lee and

---

6 See Bierens (1982) and Stinchcombe and White (1998) for a discussion on a related issue in an *i.i.d.* context.

7 This gap can be further narrowed by using the function $E(z_t^2 \mid z_{t-j}, z_{t-l})$, which may be called the bi-autoregression function of z_t at lags (j,l). An equivalent measure is the generalized third-order central cumulant function $\sigma_{j,l}^{(2,0)}(0,v)=$ cov $[z_t^2,\exp(iv_1z_{t-j}+iv_2z_{t-l})]$, where $v=(v_1,v_2)$. This is essentially a generalization of bispectral analysis and still avoids the curse of dimensionality.

Hansen, 1994, Lumsdaine, 1996). Put $\hat{z}_t = \hat{h}_t^{-1/2}\hat{\varepsilon}_t, \hat{h}_t = h_t(\hat{\theta})$, and $\hat{\varepsilon}_t = Y_t - \mu_t(\hat{\theta})$. We can estimate the generalized spectral derivative, $f^{(0,2,0)}(\omega,0,v)$, by using the following kernel estimator:

$$\hat{f}^{(0,2,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{1/2}k(j/p)\hat{\sigma}_j^{(2,0)}(0,v)e^{-\mathbf{i}j\omega}, \omega \in -\pi,\pi], v \in \mathrm{R}, \tag{3.11}$$

where

$$\hat{\sigma}_j^{(2,0)}(0,v) = -\frac{1}{T-|j|}\sum_{t=j+1}^{T}\left(\hat{z}_t^2 - 1\right)[e^{\mathbf{i}v\hat{z}_{t-j}} - \hat{\phi}_j(v)],$$

and

$$\hat{\phi}_j(v) = \frac{1}{T-|j|}\sum_{t=|j|+1}^{T}e^{\mathbf{i}v\hat{z}_{t-|j|}}.$$

Here, $p \equiv p(T)$ is a bandwidth and $k:\mathrm{R} \to -1,1]$ is a symmetric kernel. Examples of $k(\cdot)$ include the Bartlett, Daniell, Parzen, and Quadratic spectral kernels (e.g., Priestley, 1981, p.442). The factor $(1-|j|/T)^{1/2}$ is a finite-sample correction. This could be replaced by unity. Under certain conditions, $\hat{f}^{(0,2,0)}(\omega,0,v)$ is consistent for $f^{(0,2,0)}(\omega,0,v)$, as shown in Theorem 2.

On the contrary, the flat generalized spectral derivative $f_0^{(0,2,0)}(\omega,0,v)$ can be consistently estimated by using

$$\hat{f}_0^{(0,2,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\hat{\sigma}_0^{(2,0)}(0,v), \omega \in -\pi,\pi], v \in \mathrm{R}. \tag{3.12}$$

Our test is based on the quadratic form comparing (3.11) with (3.12):

$$\begin{aligned}\hat{Q} &\equiv \pi\int\int_{-\pi}^{\pi}\left|\hat{f}^{(0,2,0)}(\omega,0,v) - \hat{f}_0^{(0,2,0)}(\omega,0,v)\right|^2 d\omega dW(v)\\ &= \sum_{j=1}^{T-1}k^2(j/p)(1-j/T)\int\left|\hat{\sigma}_j^{(2,0)}(0,v)\right|^2 dW(v),\end{aligned} \tag{3.13}$$

where $W:\mathrm{R} \to \mathrm{R}^+$ is a nondecreasing weighting function that weighs sets symmetric about zero equally and the unspecified integrals are taken over the support of $W(\cdot)$. Examples of $W(\cdot)$ include the CDF of any symmetric probability distribution, either discrete or continuous. Note that the second equality follows from Parseval's identity.

3.2.1. Test Statistic Under *i.i.d.* Innovations

In many applications, practitioners often assume that the innovation $\{z_t\}$ is *i.i.d.* $(0,1)$. When this assumption is indeed valid, it is better to exploit the impact of *i.i.d.* properties of the innovations to have better finite sample performance. Our test statistic in (3.14) is still applicable. However, its finite sample performance may not be as satisfactory as when we would exploit the impact of *i.i.d.* properties of the innovations. When $\{z_t\}$: *i.i.d.* $(0,1)$, we can simplify our test statistic as follows:

$$\hat{M}^o(p) = \left[\sum_{j=1}^{T-1}k^2(j/p)(T-j)\int\left|\hat{\sigma}_j^{(2,0)}(0,v)\right|^2 dW(v) - \hat{C}^o(p)\right] / \sqrt{\hat{D}^o(p)}, \tag{3.14}$$

and the centering and scaling constants

$$\hat{C}^{o}(p) = \left(T^{-1}\sum_{t=1}^{T}\hat{z}_t^4 - 1 \right) \int \hat{\sigma}_0(v,-v)dW(v) \sum_{j=1}^{T-1} k^2(j/p),$$

$$\hat{D}^{o}(p) = 2\left(T^{-1}\sum_{t=1}^{T}\hat{z}_t^4 - 1 \right)^2 \int \left|\hat{\sigma}_0(u,v)\right|^2 dW(u)dW(v) \sum_{j=1}^{T-2} k^4(j/p),$$

with $\hat{\sigma}_0(u,v) = \hat{\phi}(u+v) - \hat{\phi}(u)\hat{\phi}(v), \hat{\phi}(v) = n^{-1}\sum_{t=1}^{n} e^{\mathrm{i}v\hat{z}_t}$. The *i.i.d.* properties of $\{z_t\}$ have been exploited to derive $\hat{C}^{o}(p)$ and $\hat{D}^{o}(p)$. The centering and scaling factors $\hat{C}^{o}(p)$ and $\hat{D}^{o}(p)$ are approximately the mean and variance of the quadratic form $T\hat{Q}$ in (3.13). Our asymptotic theory allows for both discrete and continuous weighting functions $W(\cdot)$, which weigh sets symmetric around zero equally. A continuous weighting function for $W(\cdot)$ will ensure good power for $\hat{M}^{o}(p)$, but there is a trade-off between computational cost and power when choosing a discrete or continuous weighting function $W(\cdot)$. The power of $\hat{M}^{o}(p)$ will be ensured if sufficiently fine grid points are used.

3.2.2. Test Statistic Under Non-*i.i.d.* Innovations

The volatility model with *i.i.d.* innovations $\{z_t\}$ in (2.1) is called a strong form volatility model in the literature (Drost and Nijman, 1993). However, $h_t(\cdot)$ may be correctly specified while the innovation $\{z_t\}$ displays higher order dependence, such as time-varying skewness and kurtosis. Indeed, Drost and Nijman (1993) showed that even if the innovation $\{z_t\}$ is *i.i.d.* at a certain sample frequency, when aggregated to a lower sample frequency, it will become serially dependent even if it an *m.d.s.* A volatility model where $\{z_t\}$ is not *i.i.d.* is called a semi-strong or weak form volatility model, where the innovation process is an *m.d.s.* or serially correlated (white noise), respectively. Recent studies (e.g., Gallant et al., 1991; Hansen, 1994; Harvey and Siddique, 1999 & 2000; Jondeau and Rockinger 2003) have found that the conditional skewness and kurtosis of asset returns are time-varying. Indeed, financial time series are characterized by heavy tailed non-Gaussian distributions of unknown form and these non-Gaussian features depend on market conditions. For our tests, it is also important to take into account the impact of other higher order time-varying conditional moments that may be displayed in the form of heavy tails and jumps, for example. Tests assuming *i.i.d.* innovations for $\{z_t\}$ will not be robust to time-varying conditional moments. They will have incorrect sizes (Type I errors); in particular, they may be likely to incorrectly reject correct GARCH models with time-varying higher order moments. Thus, it is highly desirable to develop tests robust to the higher order moments dynamics of unknown form. To our knowledge, no such test for volatility models exists in the literature. All existing tests assume *i.i.d.* innovations. Here, we therefore provide a test robust to time-varying higher order conditional moments of unknown form.

Our test statistic is given as follows:

$$\hat{M}(p) = \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int \left|\hat{\sigma}_j^{(2,0)}(0,v)\right|^2 dW(v) - \hat{C}(p) \right] / \sqrt{\hat{D}(p)}, \tag{3.15}$$

and

$$\hat{C}(p) = \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=j+1}^{T} \left(\hat{z}_t^4 - 1\right) \int \left|\hat{\psi}_{t-j}(v)\right|^2 dW(v),$$

$$\hat{D}(p)=2\sum_{j=1}^{T-2}\sum_{l=1}^{T-2}k^2(j/p)k^2(l/p)\iint\left|\frac{1}{T-\max(j,l)}\sum_{t=\max(j,l)+1}^{T}\left(\hat{z}_t^4-1\right)\hat{\psi}_{t-j}(u)\hat{\psi}_{t-l}^*(v)\right|^2 dW(u)dW(v),$$

and $\hat{\psi}_t(u)=e^{\mathrm{i}u\hat{z}_t}-T^{-1}\sum_{t=1}^{T}e^{\mathrm{i}u\hat{z}_t}$. Now, both centering and scaling factors are more complicated than $\hat{C}^o(p)$ and $\hat{D}^o(p)$ in the *i.i.d.* case. They have exploited the implication of a correct volatility model. At the same time, they have taken into account the impact of time-varying higher order moments of unknown form in $\{z_t\}$, such as time-varying skewness and kurtosis. This ensures a correct size for $\hat{M}^o(p)$ asymptotically. They have also incorporated the effects of time-varying higher order conditional moments (e.g., higher order conditional skewness and kurtosis). As a consequence, they are robust to non-*i.i.d.* innovations.

4. Asymptotic Distribution

The derivation of the asymptotic distribution of $\hat{M}(p)$ is much more challenging under H_0 than under *i.i.d.* because $\{z_t\}$ is not necessarily *i.i.d.* under H_0 and thus we need to take into account the possible impact of the time-varying higher order conditional moments of $\{z_t\}$. To derive the null asymptotic distribution of the test statistics $\hat{M}(p)$, we first provide some regularity conditions.

In the following, Θ is an open subset of Euclidean space R, $(\Omega, \mathrm{F}_t, P)$ is a probability space, and $I_{t-1}=\sigma(Y_{\tau-1},\varepsilon_{\tau-1},X_\tau,\tau\le t,\tau\in\mathrm{Z})$, a σ-field generated by all past observations up to $t-1$, is a sub-σ-field of F_t $(I_{t-1}\subset \mathrm{F}_t)$.

Assumption A.1: $\{Y_t\}$ is a strictly stationary process such that $Y_t=\mu_t+h_t^{1/2}z_t$, where $\mu_t\equiv E(Y_t\mid I_{t-1}), h_t\equiv \operatorname{var}(Y_t\mid I_{t-1})$, and I_{t-1} is an information set available at time $t-1$ (a σ-field generated by all past observations up to $t-1$) that may contain lagged dependent variables $\{Y_{t-j},\ j>0\}$, lagged shocks $\{\varepsilon_{t-j}\equiv h_{t-j}^{1/2}z_{t-j},\ j>0\}$, as well as current and lagged exogenous variables $\{X_{t-j},\ j\ge 0\}$.

Assumption A.2: $\mu(I_{t-1},\theta)$ and $h(I_{t-1},\theta)$ are parametric models for μ_t and h_t, where $\theta\in\Theta$ is a finite-dimensional parameter. (a) $\mu(\cdot,\theta)$ and $h(\cdot,\theta)$ are measurable with respect to I_{t-1} for each $\theta\in\Theta$; (b) with probability one, $\mu(I_{t-1},\cdot)$ and $h(I_{t-1},\cdot)$ are twice differentiable with respect to $\theta\in\Theta$; (c) for some $\nu>1$,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{n}E\sup_{\theta\in\Theta}\left\|h(I_{t-1},\theta)^{-1}\frac{\partial}{\partial\theta}h(I_{t-1},\theta)\right\|^{\max 4(\nu,2)}C,$$

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\sup_{\theta\in\Theta}\left\|h(I_{t-1},\theta)^{-\frac{1}{2}}\frac{\partial}{\partial\theta}\mu(I_{t-1},\theta)\right\|^{4}C,$$

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\sup_{\theta\in\Theta}\left\|h(I_{t-1},\theta)^{-\frac{1}{2}}\frac{\partial^2}{\partial\theta\partial\theta'}\mu(I_{t-1},\theta)\right\|^{4}C,$$

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\sup_{\theta\in\Theta}\left\|z_t h(I_{t-1},\theta)^{-1}\frac{\partial^2}{\partial\theta\partial\theta'}h(I_{t-1},\theta)\right\|^{4}C,$$

and $E\sup_{\theta\in\Theta}[z_t(\theta)]^{\max 4(\nu,2)}\le C$, where $z_t(\theta)=[Y_t-\mu_t(I_{t-1},\theta)]/\sqrt{h(I_{t-1},\theta)}$.

Assumption A.3: Let $I_t^{\dagger}$ be the observed information set available at period t that may contain some assumed initial values. Then,

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left\{E\sup_{\theta\in\Theta}\left[\frac{\mu(I_{t-1}^{\dagger},\theta)-\mu(I_{t-1},\theta)}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right]^{4}\right\}^{\frac{1}{2}} \quad C,$$

and

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left\{E\sup_{\theta\in\Theta}\left[\frac{\sqrt{h(I_{t-1}^{\dagger},\theta)}-\sqrt{h(I_{t-1},\theta)}}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right]^{8}\right\}^{\frac{1}{2}} \quad C.$$

Assumption A.4: $\hat{\theta}-\theta_0 = O_P(T^{-1/2})$, where $\theta_0 \equiv p\lim(\hat{\theta}) \in \Theta$.

Assumption A.5: The process $\{z_t, \frac{\partial}{\partial\theta'}\mu(I_{t-1},\theta_0), \frac{\partial}{\partial\theta'}h(I_{t-1},\theta_0)\}'$ is a strictly stationary α -mixing process with the α -mixing coefficient satisfying $\sum_{j=-\infty}^{\infty}\alpha(j)^{\frac{\nu-1}{\nu}} \le C$, where $\nu > 1$ is as in Assumption A.2.

Assumption A.6: The kernel $k : \mathrm{R} \to -1,1]$ is symmetric about 0, and is continuous at 0 and all points except a finite number of points, with $k(0)=1, \int_0^{\infty}k^2(z)dz < \infty$, and $|k(z)| \; C|z|^{-b}$ as $z\to\infty$ for some $b > \frac{1}{2}$.

Assumption A.7: $W : \mathrm{R} \to \mathrm{R}^{+}$ is nondegenerate, nondecreasing, and weighs sets symmetric about zero equally, with $\int_{-\infty}^{\infty}v^4 dW(v) < \infty$.

Assumption A.8: For a sufficiently large integer q, there exists a strictly stationary process $\{z_{q,t}\}$ measurable with respect to σ -field F_t such that (a) as $q\to\infty$, $z_{q,t}$ is independent of $\{z_{t-q-1}, z_{t-q-2}, \cdots\}$ for each t, $E(z_{q,t} \mid I_{t-1}) = 0$ *a.s.*, I_{t-1} is a sub-σ -field of F_t, and $E(z_{q,t}^2) = \sigma_q^2$ *a.s.*; (b) $E\left(z_t - z_{q,t}\right)^4 \; Cq^{-2\kappa}$ for some constant $\kappa \ge 1$; (c) $\sigma_q^2 \to 1$ as $q\to\infty$, and $E(z_{q,t}^2) \; \le C$ for all large q.

Assumption A.1 is a regularity condition on the data-generating process (DGP) $\{Y_t\}$. Note that $\{Y_t\}$ may not be covariance-stationary. An example is the IGARCH process, which is strictly stationary but not weakly stationary. Assumption A.2 is a standard regularity condition on the conditional mean model $\mu(I_{t-1},\theta)$ and the conditional variance model $h(I_{t-1},\theta)$. We allow for $\mu(I_{t-1},\theta)$ and $h(I_{t-1},\theta)$ to depend on the entire past history I_{t-1}, rather than a vector with a fixed dimension. This is a distinct feature from some existing nonparametric tests for conditional variance models (e.g., Hsiao and Li, 2001).

Assumption A.3 is a condition on the truncation of information set I_{t-1}, which usually contains information dating back to the remote past and so may not be observable. Because of the truncation, one may have to assume some initial values when estimating the volatility model $h(I_{t-1},\theta)$. Assumption A.3 ensures that the use of initial values, if any, does not affect the limit distribution of $\hat{M}(p)$. For instance, consider the ARMA(1,1)-GARCH(1,1) model:

$$\begin{cases} Y_t = \mu_t(\theta) + \varepsilon_t, \\ \mu_t(\theta) = \alpha Y_{t-1} + \beta \varepsilon_{t-1}, \\ \varepsilon_t = h_t(\theta)^{1/2} z_t, \\ h_t(\theta) = \gamma + \delta h_{t-1}(\theta) + \tau \varepsilon_{t-1}^2, \end{cases}$$

where $|\alpha| \le \bar{\alpha} < 1$ and $|\beta| \le \bar{\beta} < \infty$. Here, $I_{t-1} = \sigma(Y_{t-1}, Y_{t-2,} \cdots)$ but $I_{t-1}^{\dagger} = \sigma(Y_{t-1}, Y_{t-2}, \cdots, Y_1, \hat{\varepsilon}_0)$, where $\hat{\varepsilon}_0$ is an initial value assumed for ε_0. By recursive substitution, we have

$$E\sum_{t=1}^{T} \sup_{\theta\in\Theta} \left| \mu(I_{t-1}^{\dagger}, \theta) - \mu(I_{t-1}, \theta) \right| = E\sum_{t=1}^{T} \sup_{\theta\in\Theta} \left| \beta \sum_{j=t-1}^{\infty} \alpha^j \varepsilon_{t-j-1} - \beta\alpha^{t-1}\hat{\varepsilon}_0 \right|$$

$$\bar{\beta}\sum_{t=1}^{T} E \sup_{\alpha} \left| \alpha^{t-1} \left(\sum_{l=0}^{\infty} \alpha^l \varepsilon_{-l} - \hat{\varepsilon}_0 \right) \right|$$

$$2\bar{\beta}\sum_{t=1}^{T} |\bar{\alpha}|^{t-1} \left[E|\varepsilon_0| \sum_{l=0}^{\infty} \bar{\alpha}^l + E|\hat{\varepsilon}_0| \right] C.$$

We can obtain a similar condition for $h(\cdot, \theta)$ for a GARCH(1,1) model.

Assumption A.4 requires a $\sqrt{T}$ -consistent $\hat{\theta}$, which need not be the asymptotically most efficient. It can be the conditional quasi-maximum likelihood estimator in the spirit of Wooldridge's (1990 & 1991) robust modified moment-based tests for the conditional mean and variance specifications. Assumption A.5 imposes some temporal dependence conditions on the related processes. For more discussion on the mixing conditions, see White (1999).

Assumption A.6 is a regularity condition on the kernel $k(\cdot)$. It includes all commonly used kernels (Priestley, 1981, p.446). For kernels with bounded support such as the Bartlett and Parzen kernels, $b = \infty$. For the Daniell kernel, $b = 1$, and for the Quadratic-spectral kernel, $b = 2$. These kernels have unbounded support. As a consequence, all $T-1$ lags contained in the sample are used to construct the test statistics $\hat{M}(p)$. Assumption A.7 is a condition on the weighting function $W(\cdot)$ for transform parameter v. The CDF of any symmetric continuous distribution with finite variance satisfies this condition. Finally, Assumption A.8 is required only under H_0 It implies ergodicity for innovations $\{z_t\}$. It holds trivially when $\{z_t\}$ is a q-dependent process with an arbitrary but finite order. It also covers many non-Markovian processes for $\{z_t\}$.

We now state the main result of this section.

Theorem 1: *Suppose Assumptions A.1–A.8 hold, and* $p = cT^{\lambda}$ *for* $\lambda \in (0, (2b-1)/(4b-1))$ *and* $c \in (0, \infty)$. *(i) If* $\{z_t\}$ *is* $i.i.d.(0,1)$, *then* $\hat{M}^o(p) \xrightarrow{d} N(0,1)$; *(ii)* $\hat{M}(p) \xrightarrow{d} N(0,1)$ *under* H_0.

An important feature of $\hat{M}(p)$ is that the use of *estimated* standardized residuals $\{\hat{z}_t\}$ rather than unobservable innovations $\{z_t\}$ does not affect the limit distribution of $\hat{M}(p)$. One can proceed as if the true parameter value θ_0 is known and is equal to $\hat{\theta}$. This follows because the estimator $\hat{\theta}$ converges to θ_0 at the parametric rate $T^{-\frac{1}{2}}$, which is faster than the convergence rate of the nonparametric estimator $\hat{f}^{(0,2,0)}(\omega, 0, v)$. Consequently, the limit distribution of $\hat{M}(p)$ is solely determined by $\hat{f}^{(0,2,0)}(\omega, 0, v)$, and replacing $\hat{\theta}$ by θ_0 does not affect it. This delivers a convenient procedure, because it does not require any specific estimation method and one does not need to worry about the impact of parameter estimation uncertainty. Of course, parameter estimation uncertainty in $\hat{\theta}$ may still have a nontrivial impact on the small sample distribution of $\hat{M}(p)$. In this case, one may use a bootstrap procedure similar to that of Hansen (1996) to obtain more accurate sizes of the tests in small samples.

5. Asymptotic Power

Our tests are derived without assuming any alternative volatility models. To gain insight into the nature of the alternatives that our tests are able to detect, we now examine the asymptotic behavior of $\hat{M}(p)$ under the alternative to H_0.

Theorem 2: *Suppose Assumptions A.1–A.8 hold, and* $p = cT^{\lambda}$ *for* $\lambda \in (0, \frac{1}{2})$ *and* $c \in (0, \infty)$. *Then, (i)*

$$\frac{p^{1/2}}{T}\hat{M}^{o}(p) \xrightarrow{p} \left[2D^{o}\int_0^{\infty}k^4(z)dz\right]^{-1/2}\int\left|f^{(0,2,0)}(\omega,0,v) - f_0^{(0,2,0)}(\omega,0,v)\right|^2 d\omega dW(v)$$

$$= \left[2D^{o}\int_0^{\infty}k^4(z)dz\right]^{-1/2}\sum_{j=1}^{\infty}\int\left|\sigma_j^{(2,0)}(0,v)\right|^2 dW(v),$$

where $D^{o} = [E(z_t^4) - 1]^2\int|\sigma_0(u,v)|^2 dW(u)dW(v)$.

(ii)

$$\frac{p^{1/2}}{T}\hat{M}(p) \xrightarrow{p} \left[2D\int_0^{\infty}k^4(z)dz\right]^{-1/2}\int\left|f^{(0,2,0)}(\omega,0,v) - f_0^{(0,2,0)}(\omega,0,v)\right|^2 d\omega dW(v)$$

$$= \left[2D\int_0^{\infty}k^4(z)dz\right]^{-1/2}\sum_{j=1}^{\infty}\int\left|\sigma_j^{(2,0)}(0,v)\right|^2 dW(v),$$

where $D = [E(z_t^4) - 1]^2\int\left|f(\omega,u,v) - f_0(\omega,u,v)\right|^2 d\omega dW(u)dW(v)$.

Consequently, for any sequence of nonstochastic constants, $\{C_T = o(T / p^{1/2})\}$,

$$\lim_{T\to\infty}\Pr\left[\hat{M}(p) > C_T\right] = 1 \text{and} \lim_{T\to\infty}\Pr\left[\hat{M}^{o}(p) > C_T\right] = 1$$

whenever $E(z_t^2 - 1 | z_{t-j})$ *is a measurable function of* z_{t-j} *for some* $j > 0$.

We thus expect that $\hat{M}(p)$ has relatively omnibus power against a wide variety of linear and nonlinear alternatives with an unknown lag structure, as is confirmed in our simulation below. It should be emphasized that the omnibus power property does not mean that the proposed tests are more powerful than any other existing tests against *every* alternative. In fact, just because $\hat{M}(p)$ has to take care of a wide range of possible misspecifications, it may be less powerful compared with specific alternatives than a parametric test. Nevertheless, the main advantage of our omnibus test, which is not shared by any other parametric test, is that $\hat{M}(p)$ can eventually detect all possible model misspecifications that render the autoregression function $E(z_t^2 - 1 | z_{t-j})$ nonzero at any lag $j > 0$. This avoids the blindness of searching for different alternatives when one has no prior information.

Existing tests for $h(I_{t-1}, \theta)$ only consider a fixed finite number of lag orders. They can easily miss alternatives for which misspecification occurs at higher order lags. Of course, these tests could be used to check a large number of lags when a large sample is available. However, they may not be expected to be powerful when the number of lags is too large. Such a power loss due to the loss of a large number of degrees of freedom is not shared by our tests thanks to the role played by the kernel $k(\cdot)$. Most non-uniform kernels discount higher order lags (i.e., a higher order lag receives a smaller weight). This enhances good power against stationary processes whose serial dependence decays to zero as lag order j increases. Thus, our generalized spectral approach can check a large number of lags without losing too many degrees of freedom. This feature is not available for popular χ^2-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 considered.

Since $p\lim_{T\to\infty}\hat{M}(p) = +\infty$ whenever $\mathrm{var}\left(z_t \mid z_{t-j}\right) \neq 1$ for some lag $j > 0$, the appropriate critical values of $\hat{M}^o(p)$ and $\hat{M}(p)$ are the upper-tailed asymptotic critical values (e.g., 1.645 at the 5% level). We note that the rate conditions on lag order $p \equiv p(T)$ are different under the null hypothesis (Theorem 1) and the alternative hypothesis (Theorem 2). The tests are applicable under the more restrictive condition on p of Theorem 1.

6. Data-Driven Lag Order

A practical issue in implementing our tests is the choice of the lag order or bandwidth p. One advantage of our generalized spectral approach is that it can provide a data-driven method to choose p, which, to some extent, let data themselves speak for a proper p. Before discussing specific data-driven methods, we first justify the use of the data-driven lag order, $\hat{p}$ say. For this purpose, we impose a Lipschitz continuity condition on the kernel $k(\cdot)$. This condition rules out the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$, where $\mathbf{1}(\cdot)$ is the indicator function, but it still includes most commonly used kernels.

Assumption A.9: *For any* $x, y \in \mathrm{R}$, $|k(x) - k(y)| \leq C|x - y|$ *for some constant* C.

Theorem 3: *Suppose Assumptions A.1–A.9 hold, and* $\hat{p}$ *is a data-driven bandwidth such that* $\hat{p}/p = 1 + O_P(p^{-(\frac{3}{2}\beta-1)})$ *for some* $\beta > (2b - \frac{1}{2})/(2b-1)$, *where* b *is as in Assumption A.5, and p is a nonstochastic bandwidth with* $p = cT^{\lambda}$ *for* $\lambda \in (0, (2b-1)/(4b-1))$ *and* $c \in (0, \infty)$. *Then (i) If* $\{z_t\}$ *is i.i.d.* (0,1), *then* $\hat{M}^o(\hat{p}) - \hat{M}^o(p) \xrightarrow{p} 0$ *and* $\hat{M}^o(\hat{p}) \xrightarrow{d} N(0,1)$; *(i) under* H_0, $\hat{M}(\hat{p}) - \hat{M}(p) \xrightarrow{p} 0$ *and* $\hat{M}(\hat{p}) \xrightarrow{d} N(0,1)$.

Thus, as long as $\hat{p}$ converges to p sufficiently fast, the use of $\hat{p}$ rather than p does not affect the limit distribution of $\hat{M}(\hat{p})$. This is an additional "nuisance parameter-free" property.

Theorem 3 allows for a wide range of admissible rates for $\hat{p}$. One plausible choice of $\hat{p}$ is the nonparametric plug-in method considered in Hong (1999). This minimizes an asymptotic integrated mean squared error (IMSE) criterion for the estimator $\hat{f}^{(0,2,0)}(\omega, 0, v)$. Nonparametric plug-in methods are common in the literature (e.g., Newey and West, 1994, Silverman, 1986). Consider some "pilot" generalized spectral derivative estimators based on a preliminary bandwidth $\bar{p}$:

$$\bar{f}^{(0,2,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{\frac{1}{2}}\bar{k}(j/\bar{p})\hat{\sigma}_j^{(2,0)}(0, v)e^{-\mathrm{i}j\omega}, \tag{6.1}$$

$$\bar{f}^{(q,2,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=1-T}^{T-1}(1-|j|/T)^{\frac{1}{2}}\bar{k}(j/\bar{p})\hat{\sigma}_j^{(2,0)}(0,v)\,|j|^q\, e^{-\mathbf{i}j\omega}, \tag{6.2}$$

where the kernel $\bar{k}(\cdot)$ need not be the same as the kernel $k(\cdot)$ used in (3.14). For example, $\bar{k}(\cdot)$ can be the Bartlett kernel, while $k(\cdot)$ is the Daniell kernel. Note that $\bar{f}^{(0,2,0)}(\omega,0,v)$ is an estimator of $f^{(0,2,0)}(\omega,0,v)$ and $f^{(q,2,0)}(\omega,0,v)$ is an estimator of the generalized spectral derivative

$$f^{(q,2,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j^{(2,0)}(0,v)\,|j|^q\, e^{-\mathbf{i}j\omega}. \tag{6.3}$$

Suppose for the kernel $k(\cdot)$, there exists some $q \in (0,\infty)$ such that $0 < k^{(q)} \equiv \lim_{z\to 0}\frac{1-k(z)}{|z|^q} < \infty.$ Then, the plug-in bandwidth is defined as

$$\hat{p}_0 \equiv \hat{c}_0 T^{\frac{1}{2q+1}}, \tag{6.4}$$

where the tuning parameter estimator is

$$\hat{c}_0 \equiv \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)dz}\frac{\iint_{-\pi}^{\pi}\left|\bar{f}^{(q,2,0)}(\omega,0,v)\right|^2 d\omega dW(v)}{Re\int_{-\pi}^{\pi}\int \bar{f}^{(0,2,0)}(\omega,0,v)\,\bar{f}(\omega,v,-v)dW(v)d\omega}\right]^{\frac{1}{2q+1}}$$

$$= \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)dz}\frac{\sum_{j=1-T}^{T-1}(T-|j|)\bar{k}^2(j/\bar{p})\,|j|^{2q}\int\left|\hat{\sigma}_j^{(2,0)}(0,v)\right|^2 dW(v)}{\sum_{j=1-T}^{T-1}(T-|j|)\bar{k}^2(j/\bar{p})\hat{R}_2(j)Re\int\hat{\sigma}_j(v,-v)dW(v)}\right]^{\frac{1}{2q+1}},$$

with $\hat{R}_2(j) = T^{-1}\sum_{t=j+1}^{T}(\hat{z}_t^2-1)(\hat{z}_{t-j}^2-1).$ The second equality here follows from Parseval's identity. Note that $\hat{p}_0$ is real-valued. One can take its integer part, and the impact of integer clipping is expected to be negligible.

The data-driven $\hat{p}_0$ in (6.4) involves the choice of a preliminary bandwidth $\bar{p}$, which can be either fixed or growing with the sample size T. If $\bar{p}$ is fixed, $\hat{p}_0$ generally grows at the rate $T^{\frac{1}{2q+1}}$ under H_A, but $\hat{c}_0$ does not converge to the optimal tuning constant that minimizes the IMSE of $\hat{f}^{(0,2,0)}(\omega,0,v)$. This is analogous in spirit to a parametric plug-in method. Following Hong (1999), we can show that when $\bar{p}$ grows with T properly, the data-driven bandwidth $\hat{p}_0$ in (6.4) minimizes the asymptotic IMSE of $\hat{f}^{(0,2,0)}(\omega,0,v)$. Although the choice of $\bar{p}$ is somewhat arbitrary, we expect that it is of secondary importance. This is confirmed in our simulation below.

From a theoretical point of view, the choice of $\hat{p}$ based on the IMSE criterion may not maximize the power of the test. A more sensible alternative would be to develop a data-driven $\hat{p}$ using a power criterion, or a criterion that trades off size distortion and power loss. This will necessitate higher order asymptotic analysis and is beyond the scope of this study. We are content with the IMSE criterion here. Our simulation experience suggests that the power of our tests is relatively flat in the neighborhood of the optimal lag order that maximizes the power and that the data-driven $\hat{p}_0$ based on IMSE performs reasonably well in finite samples.

7. Monte Carlo Evidence

We now investigate the finite sample performance of our tests. Because our tests are derived without specifying an alternative, we compare them with several popular tests of similar spirit, namely the Box–Pierce–Ljung-type test for squared standardized errors and Li and Mak's (1994) test. These tests assume that the standardized innovations $\{z_t\}$ are *i.i.d*.

7.1. DGPs

We consider the following DGPs:

DGP 1 [AR(1)-GARCH(1,1)]:

$$\begin{cases} Y_t = 0.2Y_{t-1} + \varepsilon_t, \\ \varepsilon_t = \sqrt{h_t} z_t, \\ h_t = 0.2 + 0.6h_{t-1} + 0.2\varepsilon_{t-1}^2, \\ \{z_t\} : i.i.d.N(0,1). \end{cases}$$

DGP 2 [AR(1)-Threshold GARCH(1,1)]:

$$\begin{aligned} &Y_t = 0.2Y_{t-1} + \varepsilon_t, \\ &\varepsilon_t = \sqrt{h_t} z_t, \\ &h_t = 0.2 + 0.6h_{t-1} + 0.2\varepsilon_{t-1}^2 1(\varepsilon_{t-1} \ge 0) + 0.5\varepsilon_{t-1}^2 1(\varepsilon_{t-1} < 0), \\ &\{z_t\} : i.i.d.\mathrm{N}(0,1). \end{aligned}$$

DGP 3 [AR(1)-[EGARCH(1,1)]:

$$\begin{aligned} &Y_t = 0.2Y_{t-1} + \varepsilon_t, \\ &\varepsilon_t = \sqrt{h_t} z_t, \\ &\ln h_t = 0.01 + 0.9\ln h_{t-1} + 0.3(|z_{t-1}| - \sqrt{\frac{2}{\pi}}) - 0.08z_{t-1}, \\ &\{z_t\} : i.i.d.\mathrm{N}(0,1). \end{aligned}$$

We use the GAUSS Windows version random number generator to generate data with the sample sizes $T = 500$, 100, and 2000. For each data set, we first generate $T + 500$ observations and the discard the first 500 to reduce the impact of the initial values. We then use the QMLE method to estimate the following AR(1)-GARCH(1,1) model:

$$\begin{aligned} &Y_t = \alpha Y_{t-1} + \varepsilon_t, \\ &\varepsilon_t = \sqrt{h_t} z_t, \\ &h_t = \varphi + \beta h_{t-1} + \gamma\varepsilon_{t-1}^2, \\ &\{z_t\} : i.i.d.(0,1). \end{aligned}$$

Under DGP 1, the AR(1)-GARCH(1,1) model is correctly specified. This allows us to examine the sizes of the tests in finite samples. The AR(1)-GARCH(1,1) model is misspecified under DGPs 2 and 3. These two DGPs are the threshold GARCH(1,1) and EGARCH(1,1) processes, respectively, which have nonlinear volatility dynamics. they thus allow us to examine the power of the tests to detect the neglected nonlinearity in volatility dynamics. For DGP 1, we generate 1000 data sets for each T. For DGPs 2 and 3, we generate 500 data sets for each T.

To compute $\hat{M}(\hat{p}_0)$, we use the N(1,0) CDF truncated on $[-3,3]$ for the weighting function $W(\cdot)$. We use the Bartlett kernel $k_B(z)=(1-|z|)\mathbf{1}(|z|\le 1)$ for $k(\cdot)$, which has bounded support and is computationally efficient. Our simulation experience suggests that the choices of $W(\cdot)$ and $k(\cdot)$ have little impact on the size and power of our tests.[8] We choose a data-driven $\hat{p}_0$ via the plug-in method in (6.4), with the Bartlett kernel for $\bar{k}(\cdot)$ used in the preliminary generalized spectral derivative estimators in (6.1) and (6.2). To examine the impact of the choice of preliminary bandwidth $\bar{p}$, we consider $\bar{p}=5,10,15,20,25,$ and 30.

7.2. Monte Carlo Evidence

Table 1 reports the empirical rejection rates of the tests under H_0 at the 10% and 5% significance levels, using asymptotic theory. Under DGP 1 (the null model), the generalized spectral test $\hat{M}^o(\hat{p}_0)$ underrejects H_0, but not excessively. The sizes increase as the sample size T increases, and slightly overreject at the 5% significance level with $T=2000$. The $\hat{M}(\hat{p}_0)$ test also underrejects H_0, but the sizes improve as the sample size T increases. The $\hat{M}(\hat{p}_0)$ test has better sizes than $\hat{M}^o(\hat{p}_0)$ as is expected. A larger preliminary lag order $\bar{p}$ tends to give a better size for both $\hat{M}^o(\hat{p}_0)$ and $\hat{M}(\hat{p}_0)$ tests at the 10% level. However, at the 5% level, both tests seem to be robust to the choice of $\bar{p}$. Overall, the sizes of both $\hat{M}^o(\hat{p}_0)$ and $\hat{M}(\hat{p}_0)$ are reasonable in finite samples.

Next, we turn to the sizes of $\mathrm{BP}(p)$ and $\mathrm{LM}(p)$. The former is not valid asymptotically but we include it for comparison purposes. The latter is a valid test under DGP 1. Both tests underreject, but have better sizes than the $\hat{M}(\hat{p}_0)$ test with $T=500$. However, their sizes do not improve when the sample size T increases and become worse than $\hat{M}(\hat{p}_0)$ with $T=1000$ and 2000.

Table 1. Empirical Sizes of the Tests at the 10% and 5% Significance Levels

	DGP 1: AR(1)- GARCH(1,1)-i.i.d.N(0,1)							
	$M(p)$		$M^o(p)$		$\mathrm{BP}_2(p)$		$\mathrm{LM}(p)$	
p	10%	5%	10%	5%	10%	5%	10%	5%
				$T=500$				
5	3.5	1.8	6.8	4.2	4.7	2.5	3.9	2.0
10	3.5	1.8	6.8	4.2	6.4	2.2	5.4	2.1
15	3.5	1.8	6.8	4.2	6.1	3.1	5.3	2.6

8 We also used the Parzen kernel (not reported). The test statistics are rather similar to those based on the Bartlett kernel in most cases.

Table 1. *Cont.*

	DGP 1: AR(1)- GARCH(1,1)-i.i.d.N(0,1)							
	$M(p)$		$M^{o}(p)$		$\mathrm{BP}_2(p)$		$\mathrm{LM}(p)$	
p	10%	5%	10%	5%	10%	5%	10%	5%
				$T = 500$				
20	3.5	1.8	6.9	4.5	7.2	4.8	5.9	3.6
25	3.4	1.7	7.2	4.3	8.9	5.1	7.3	4.3
30	3.5	1.7	7.5	4.1	10.1	5.5	7.4	4.3
				$T = 1000$				
5	3.6	2.3	6.4	3.3	3.7	2.0	3.1	1.6
10	3.6	2.3	6.4	3.3	4.6	2.5	4.3	2.2
15	3.6	2.3	6.4	3.3	6.7	3.2	6.1	2.9
20	4.2	2.3	6.4	3.3	6.8	3.3	5.9	2.9
25	4.0	2.2	7.1	3.5	7.7	3.4	6.2	3.0
30	6.4	2.0	7.2	4.0	7.8	1.6	6.3	3.0
				$T = 2000$				
5	6.4	4.2	8.1	5.6	4.6	2.2	3.7	1.7
10	6.4	4.2	8.1	5.6	6.0	2.7	5.2	2.5
15	6.4	4.2	8.1	5.6	6.5	2.9	5.9	2.9
20	6.4	4.2	8.1	5.6	6.6	3.3	6.4	3.3
25	6.6	4.3	8.4	5.8	7.3	3.4	7.1	2.9
30	7.1	4.2	9.6	5.7	6.7	4.0	6.3	3.6

Notes: (i) 1000 iterations; (ii) $\hat{M}(\hat{p}), \hat{M}^{o}(\hat{p})$, generalized spectral tests derived under non-i.i.d. and *i.i.d.* innovations, respectively; $BP(p)$, Box and Pierce's (1970) type test; $LM(p)$, Li and Mak's (1994) test; (iii) The Bartlett kernel is used for both $\hat{M}(\hat{p})$ and $\hat{M}^{o}(\hat{p})$; (iv) DGP 1: $Y_t = 0.2Y_{t-1} + \varepsilon_t, \varepsilon_t = \sqrt{h_t} z_t$, $h_t = 0.2 + 0.6h_{t-1} + 0.2\varepsilon_{t-1}^2$, $z_t \sim i.i.d.\, N(0,1)$.

Table 2 reports the level-corrected power of the tests at the 10% and 5% levels under DGPs 1 and 2, based on the empirical critical values. The empirical critical values are obtained under DGP 1. Under DGP 2 (AR(1)-Threshold GARCH(1,1)) and DGP 3 (AR(1)-EGARCH(1,0)), $\hat{M}^{o}(\hat{p}_0)$ and $\hat{M}(\hat{p}_0)$ are powerful. Their power increases as the sample size T increases. The power of $\hat{M}^{o}(\hat{p}_0)$ and $\hat{M}(\hat{p}_0)$ is also relatively robust to the choice of the preliminary lag order $\overline{p}$. In sharp contrast, $\mathrm{BP}(p)$ and $\mathrm{LM}(p)$ have no power compared with DGPs 2 and 3. Their empirical rejection rates are close to the significance levels. These results confirm our theory that the proposed new tests can effectively detect neglected nonlinearity in volatility dynamics, a crucial feature that existing tests do not share.

Table 2. Size-Corrected Powers of the Tests at the 10% and 5% Significance Levels

	DGP 2: AR(1)-Threshold GARCH(1,1)								DGP 3: AR(1)-EGARCH(1,0)							
	$M(\hat{p}_0)$		$M^o(\hat{p}_0)$		BP (p)		LM (p)		$M(\hat{p}_0)$		$M^o(\hat{p}_0)$		BP (p)		LM (p)	
p	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
			$T = 500$								$T = 500$					
5	54.6	40.8	54.6	43.2	12.2	6.6	10.8	6.2	29.6	21.4	30.8	23.0	10.4	5.0	10.0	5.0
	54.6	40.8	54.6	43.2	13.2	7.8	13.2	8.0	29.6	21.4	30.8	23.0	11.2	6.2	10.6	6.0
15	54.6	40.8	54.6	43.2	13.0	7.4	12.4	7.4	29.6	21.4	30.8	23.0	10.0	6.2	10.0	6.6
20	54.8	40.8	54.6	42.8	12.4	6.8	11.4	6.8	29.8	21.8	30.8	22.4	11.0	5.2	11.0	5.2
25	54.0	39.2	53.8	41.6	11.8	6.6	11.8	6.4	30.6	20.8	32.2	22.4	12.0	5.2	12.0	5.0
30	49.8	39.2	50.0	41.0	11.8	6.4	11.4	6.4	31.0	20.2	31.2	21.2	9.8	5.2	10.4	5.2
			$T = 1000$								$T = 1000$					
5	84.8	76.8	85.4	78.2	14.4	6.6	14.8	6.4	54.0	43.0	56.2	43.6	11.2	5.2	11.2	5.4
10	84.8	76.8	85.4	78.2	14.0	7.4	13.0	6.6	54.0	43.0	56.2	43.6	9.2	4.8	9.0	4.2
15	84.8	76.8	85.4	78.2	13.0	7.8	12.6	7.2	54.0	43.0	56.2	43.6	10.8	4.6	10.6	4.2
20	84.8	76.8	85.4	78.2	12.4	7.4	11.2	7.6	53.4	43.0	56.2	43.6	10.0	5.4	9.8	5.4
25	84.2	76.4	84.8	77.4	13.8	7.4	13.0	7.4	52.6	42.0	55.2	43.0	10.4	4.8	10.0	5.0
30	83.8	73.6	84.6	75.8	12.2	7.2	11.4	7.2	51.6	38.6	55.0	40.6	9.0	4.8	8.8	5.2
			$T = 2000$								$T = 2000$					
5	98.8	98.0	98.8	98.0	14.4	8.4	14.2	8.8	84.4	76.4	83.6	76.8	8.6	4.2	9.8	4.8
10	98.8	98.0	98.8	98.0	13.2	7.2	12.4	6.6	84.4	76.4	83.6	76.8	8.6	5.2	8.4	5.2
15	98.8	98.0	98.8	98.0	11.4	5.6	11.2	5.6	84.4	76.4	83.6	76.8	10.2	4.8	9.8	4.6
20	98.8	98.0	98.8	98.0	13.6	6.2	12.4	6.0	84.4	76.4	83.6	76.8	8.4	3.6	8.0	3.4
25	98.8	98.0	98.8	98.0	13.6	6.4	13.2	6.4	84.2	76.6	83.8	75.8	8.0	5.0	7.8	5.0
30	98.8	97.6	98.8	97.6	12.4	6.8	12.4	6.4	83.6	75.4	83.0	74.6	8.6	4.8	8.4	4.6

Notes: (i) 500 iterations; (ii) $\hat{M}(\hat{p}_0), \hat{M}^o(\hat{p}_0)$, generalized spectral tests derived under non-i.i.d. and *i.i.d.* innovations, respectively; $BP(p)$, Box and Pierce's (1970) test; $LM(p)$, Li and Mak's (1994) test; (iii) The Bartlett kernel is used for both $M(\hat{p}_0)$ and $M^o(\hat{p}_0)$; (iv) DGP 2, $Y_t = 0.2Y_{t-1} + \varepsilon_t, \varepsilon_t = \sqrt{h_t} z_t, h_t = 0.2 + 0.6h_{t-1} + 0.2\varepsilon_{t-1}^2 1(\varepsilon_{t-1} \geq 0) + 0.5\varepsilon_{t-1}^2 1(\varepsilon_{t-1} < 0),\ z_t : i.i.d.N(0,1)$; DGP 3, $Y_t = 0.2Y_{t-1} + \varepsilon_t, \varepsilon_t = \sqrt{h_t} z_t, \ln h_t = 0.01 + 0.9 \ln h_{t-1} + 0.3(|z_{t-1}| - \sqrt{\frac{2}{\pi}}) - 0.08 z_{t-1}, z_t : i.i.d.N(0,1).$

In summary, we observe the following two points: (i) the empirical sizes of the generalized spectral tests $\hat{M}^o(\hat{p}_0)$ and $\hat{M}(\hat{p}_0)$ are smaller than the nominal levels, but they improve and become reasonable as the sample size increases; and (ii) both the $\hat{M}^o(\hat{p}_0)$ and the $\hat{M}(\hat{p}_0)$ tests are powerful at detecting neglected nonlinearity in volatility dynamics and their powers are relatively robust with respect to the choice of the preliminary lag order $\bar{p}$. In sharp contrast, $\mathrm{BP}(p)$ and $\mathrm{LM}(p)$ have little power against the neglected nonlinear volatility dynamics. These results suggest that the proposed tests can be useful diagnostic tools for practitioners when modeling volatility dynamics.

We emphasize that we only considered two nonlinear volatility dynamics in the simulation study. If a linear volatility alternative (e.g., an AR(1)-GARCH(2,2) model) is considered, the $\hat{M}^{o}(\hat{p}_0)$ and $\hat{M}(\hat{p}_0)$ tests may not be as powerful as the $\mathrm{BP}(p)$ and $\mathrm{LM}(p)$ tests in finite samples, especially when an optimal lag order p is used for the latter. Further, we also only considered *i.i.d.* standardized innovations in our simulation study, which are required to compare the $\mathrm{BP}(p)$ and $\mathrm{LM}(p)$ tests. For non-*i.i.d.* standardized innovations, $\hat{M}^{o}(\hat{p}_0)$ becomes invalid, while $\hat{M}(\hat{p}_0)$ remains asymptotically valid. Given Hong and Lee's (2005) simulation study of generalized spectral tests for time series conditional mean models, we expect $\hat{M}^{o}(\hat{p}_0)$ to strongly overreject in finite samples, while $\hat{M}(\hat{p}_0)$ tends to underreject slightly under . H_0 . with non-*i.i.d.* standardized innovations.

8. Conclusions

Volatility models have played important roles in economics and finance such as in studies of the trade-off between return and risk, volatility clustering, and volatility spillovers among financial markets or between financial sectors and real sectors. By using a second-order generalized spectral derivative approach, we propose a class of new diagnostic tests for volatility models in time series, where the dimension of the conditioning information set may be infinite and the volatility models can be linear or nonlinear. The tests can detect a wide range of model misspecifications in volatility models, while being robust to the higher order time-varying moments of unknown form (e.g., skewness and kurtosis). They can also check a large number of lags and naturally discount higher order lags, which is consistent with the stylized fact that economic or financial markets are more affected by recent past events than remote past events. No specific estimation method is required, and the tests have the appealing "nuisance parameter-free" property that parameter estimation uncertainty does not affect the limit distribution of the test statistics. Only the standardized estimated residuals are needed to carry out our tests. Further, there is no need to compute the score functions of the volatility models, as required by some existing tests. We examine the finite sample performance of the proposed tests in a simulation study, finding that they can detect the neglected nonlinear volatility dynamics that existing tests can easily miss. They are thus useful diagnostic tools for practitioners when modeling volatility dynamics.

Acknowledgments: We thank Liyuan Cui, Xia Wang, and Chengguang Zhong for their comments and suggestions.
Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix, we outline the proofs of Theorems 1–3. Detailed proofs can be found in the Supplementary Material, which is available online. Throughout this appendix, we define $M(p)$ and $M^{o}(p)$ in the same way as $\hat{M}(p)$ and $\hat{M}^{o}(p)$ in (3.14) and (3.15), with the unobservable standardized residual sample $\{z_t \equiv z_t(\theta_0)\}_{t=1}^{T}$, where $\theta_0 \equiv p\lim\hat{\theta}$, replacing the estimated standardized residual sample $\{\hat{z}_t\}_{t=1}^{T}$. Further, $C \in (1,\infty)$ denotes a generic bounded constant.

Proof of Theorem 1: To save space, we only consider $\hat{M}(p)$; the proof for $\hat{M}^{o}(p)$ is similar and simpler. It suffices to show Theorems A.1–A.3. Theorem A.1 implies that the use of the estimated standardized sample $\{\hat{z}_t\}_{t=1}^{T}$ rather than the unobservable sample $\{z_t\}_{t=1}^{T}$ does not affect the limit distribution of $\hat{M}(p)$. Theorem A.2 implies that the use of the truncated standardized disturbances $\{z_{q,t}\}_{t=1}^{T}$ rather than $\{z_t\}_{t=1}^{T}$ does not affect the limit distribution of $\hat{M}(p)$ for sufficiently large q. The assumption that $z_{q,t}$ is independent of $\{z_{t-j}\}_{j=q+1}^{\infty}$ when q is large simplifies the proof of the asymptotic normality of $\hat{M}(p)$ considerably.

Theorem A.1: *Under the conditions of Theorem 1,* $\hat{M}(p)-M(p)\xrightarrow{p}0$.

Theorem A.2: *Let* $M_q(p)$ *be defined as* $M(p)$ *with* $\{z_{q,t}\}_{t=1}^{T}$ *replacing* $\{z_t\}_{t=1}^{T}$, *where* $\{z_{q,t}\}$ *is as in Assumption A.1. Then, under the conditions of Theorem 1 and* $q=p^{1+\frac{1}{4b-2}}T^{\frac{2b}{2b-1}}$, $M_q(p)-M(p)\xrightarrow{p}0$.

Theorem A.3: *Under the conditions of Theorem 1 and* $q=p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$, $M_q(p)\xrightarrow{d}N(0,1)$.

Proof of Theorem A.1: Noting that $z_t(\theta)\equiv Y_t-\mu(I_{t-1},\theta)]/\sqrt{h(I_{t-1},\theta)}$ in (3.1), where I_t is the unobservable information set from period t to the infinite past, we write

$$
\begin{aligned}
\hat{z}_t &\equiv \frac{Y_t-\mu(I_{t-1}^{\dagger},\hat{\theta})}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}} \\
&= \frac{Y_t-\mu(I_{t-1},\hat{\theta})}{\sqrt{h(I_{t-1},\hat{\theta})}}\frac{\sqrt{h(I_{t-1},\hat{\theta})}}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}+\frac{\mu(I_{t-1},\hat{\theta})-\mu(I_{t-1}^{\dagger},\hat{\theta})}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}} \\
&= z_t(\hat{\theta})+z_t(\hat{\theta})\frac{\sqrt{h(I_{t-1},\hat{\theta})}-\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}+\frac{\mu(I_{t-1},\hat{\theta})-\mu(I_{t-1}^{\dagger},\hat{\theta})}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}.
\end{aligned}
\tag{A.1}
$$

From the mean value theorem, we have

$$
z_t(\hat{\theta})=\frac{Y_t-\mu(I_{t-1},\theta_0)}{\sqrt{h(I_{t-1},\theta_0)}}+\xi_t(\bar{\theta})'(\hat{\theta}-\theta_0)=z_t(\theta_0)+\xi_t(\bar{\theta})'(\hat{\theta}-\theta_0)
\tag{A.2}
$$

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

$$
\xi_t(\theta)\equiv\frac{\partial}{\partial\theta}z_t(\theta)=-\frac{1}{2}z_t(\theta)\frac{1}{h(I_{t-1},\theta)}\frac{\partial}{\partial\theta}h(I_{t-1},\theta)-\frac{1}{\sqrt{h(I_{t-1},\theta)}}\frac{\partial}{\partial\theta}\mu(I_{t-1},\theta).
$$

It follows from (A.1), (A.2), and Markov's inequality that

$$
\begin{aligned}
{}_{t=1}^{T}\left[\hat{z}_t-z_t(\hat{\theta})\right]^2 &= \sum_{t=1}^{T}\left[\frac{\mu(I_{t-1},\hat{\theta})-\mu(I_{t-1}^{\dagger},\hat{\theta})}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}+z_t(\hat{\theta})\frac{\sqrt{h(I_{t-1},\hat{\theta})}-\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}\right]^2 \\
&\le 2\sum_{t=1}^{T}\left[\frac{\mu(I_{t-1},\hat{\theta})-\mu(I_{t-1}^{\dagger},\hat{\theta})}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}\right]^2+2\sum_{t=1}^{T}z_t^2(\hat{\theta})\left[\frac{\sqrt{h(I_{t-1},\hat{\theta})}-\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}{\sqrt{h(I_{t-1}^{\dagger},\hat{\theta})}}\right]^2 \\
&= O_P(1),
\end{aligned}
\tag{A.3}
$$

where we have used the fact that

$$E\sum_{t=1}^{T} z_t^2(\hat\theta)\left[\frac{\sqrt{h(I_{t-1},\hat\theta)}-\sqrt{h(I_{t-1}^{\dagger},\hat\theta)}}{\sqrt{h(I_{t-1}^{\dagger},\hat\theta)}}\right]^2 \le \sum_{t=1}^{T}\left[E\sup_{\theta\in\Theta_0} z_t^4(\theta)\right]^{\frac{1}{2}}\left\{E\sup_{\theta\in\Theta_0}\left[\frac{\sqrt{h(I_{t-1},\theta)}-\sqrt{h(I_{t-1}^{\dagger},\theta)}}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right]^4\right\}^{\frac{1}{2}}$$

$$\le C\sum_{t=1}^{T}\left\{E\sup_{\theta\in\Theta_0}\left[\frac{\sqrt{h(I_{t-1},\theta)}-\sqrt{h(I_{t-1}^{\dagger},\theta)}}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right]^4\right\}^{\frac{1}{2}}$$

$$\le C^2,$$

by the Cauchy-Schwarz inequality and Assumptions A.2 and A.3. Here, the second term in the third inequality is $O_P(1)$ according to Markov's inequality and Assumptions A.1 and A.3. On the contrary, from (A.2) and Assumptions A.2–A.4, we have

$$\sum_{t=1}^{T}\left[z_t(\hat\theta)-z_t(\theta_0)\right]^2=\sum_{t=1}^{T}\left[\xi_t(\bar\theta)'(\hat\theta-\theta_0)\right]^2\le\|\sqrt{T}(\hat\theta-\theta_0)\|^2\left[T^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^2\right]$$

$$=O_P(1), \tag{A.4}$$

where we used the fact that

$$E\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^2\le\left[E\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^4\right]^{\frac{1}{2}}$$

and

$$E\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^4=E\sup_{\theta\in\Theta_0}\left\|\frac{1}{2}z_t(\theta)\frac{h'(I_{t-1},\theta)}{h(I_{t-1},\theta)}-\frac{\mu'(I_{t-1},\theta)}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right\|^4$$

$$\le 2E\sup_{\theta\in\Theta_0}\left\|\frac{1}{2}z_t(\theta)\frac{h'(I_{t-1},\theta)}{h(I_{t-1},\theta)}\right\|^4+2E\sup_{\theta\in\Theta_0}\left\|\frac{\mu'(I_{t-1},\theta)}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right\|^4$$

$$\le\frac{1}{2^3}\left[E\sup_{\theta\in\Theta_0}z_t^8(\theta)\right]^{\frac{1}{2}}\left[E\sup_{\theta\in\Theta_0}\left\|\frac{h'(I_{t-1},\theta)}{h(I_{t-1},\theta)}\right\|^8\right]^{\frac{1}{2}}+2E\sup_{\theta\in\Theta_0}\left\|\frac{\mu'(I_{t-1},\theta)}{\sqrt{h(I_{t-1}^{\dagger},\theta)}}\right\|^4$$

$$\le C,$$

where $\mu'(I_{t-1},\theta)=\frac{\partial}{\partial\theta}\mu(I_{t-1},\theta)$ and $h'(I_{t-1},\theta)=\frac{\partial}{\partial\theta}h(I_{t-1},\theta)$, given Assumption A.2. Both (A.3) and (A.4) imply

$$\sum_{t=1}^{T}\left[\hat z_t-z_t(\theta_0)\right]^2=O_P(1). \tag{A.5}$$

Put $T_j=T-|j|$. Observe that $p\to\infty, p/T\to 0, p^{-1}\sum_{j=1}^{T-1}k^r(j/p)\to\int_0^\infty k^r(z)dz$ for $r=2,4$ given Assumption A.6. To show $\hat M(p)-M(p)\xrightarrow{p}0$, it suffices to show

$$p^{-1/2}\int\sum_{j=1}^{T-1}k^2(j/p)T_j\left[|\hat{\sigma}_j^{(2,0)}(0,v)|^2-|\tilde{\sigma}_j^{(2,0)}(0,v)|^2\right]dW(v)\xrightarrow{p}0, \tag{A.6}$$

$p^{-1}[\hat{C}(p)-\tilde{C}(p)]=O_P(T^{-\frac{1}{2}})$ and $p^{-1}[\hat{D}(p)-\tilde{D}(p)]\xrightarrow{p}0$, where $\tilde{C}(p)$ and $\tilde{D}(p)$ are defined in the same way as $\hat{C}(p)$ and $\hat{D}(p)$ in (3.14), with $\{z_t\}_{t=1}^T$ replacing $\{\hat{z}_t\}_{t=1}^T$. To save space, we focus on the proof of (A.6); the proofs of $p^{-1}[\hat{C}(p)-\tilde{C}(p)]=O_P(T^{-\frac{1}{2}})$ and $p^{-1}[\hat{D}(p)-\tilde{D}(p)]\xrightarrow{p}0$ are straightforward. We note that it is necessary to obtain the convergence rate $O_P(T^{-\frac{1}{2}})$ for $p^{-1}[\hat{C}(p)-\tilde{C}(p)]$ to ensure that replacing $\hat{C}(p)$ with $\tilde{C}(p)$ has an asymptotically negligible impact given $p/T\to 0$.

To show (A.6), we first decompose

$$\int\sum_{j=1}^{T-1}k^2(j/p)T_j\left[|\hat{\sigma}_j^{(2,0)}(0,v)|^2-|\tilde{\sigma}_j^{(2,0)}(0,v)|^2\right]dW(v)=\hat{A}_1+2Re(\hat{A}_2), \tag{A.7}$$

where

$$\hat{A}_1=\int\sum_{j=1}^{T-1}k^2(j/p)T_j\left|\hat{\sigma}_j^{(2,0)}(0,v)-\tilde{\sigma}_j^{(2,0)}(0,v)\right|^2dW(v), \tag{A.8}$$

$$\hat{A}_2=\int\sum_{j=1}^{T-1}k^2(j/p)T_j\left[\hat{\sigma}_j^{(2,0)}(0,v)-\tilde{\sigma}_j^{(2,0)}(0,v)\right]\tilde{\sigma}_j^{(2,0)}(0,v)^*dW(v), \tag{A.9}$$

where $Re(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\tilde{\sigma}_j^{(2,0)}(0,v)^*$ is the complex conjugate of $\tilde{\sigma}_j^{(2,0)}(0,v)$. Then, (A.6) follows from Propositions A.1 and A.2 and $p\to\infty$ as $T\to\infty$.

Proposition A.1: *Under the conditions of Theorem 1,* $\hat{A}_1=O_P(1)$.

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

Proof of Proposition A.1: Put $\hat{\delta}_t(v)\equiv e^{\mathrm{i}v\hat{z}_t}-e^{\mathrm{i}vz_t}$ and $\psi_t(v)\equiv e^{\mathrm{i}vz_t}-\phi(v)$, where, as before, $\phi(v)\equiv E(e^{\mathrm{i}vz_t})$. Then, straightforward algebra yields that for $j>0$,

$$\begin{aligned}
-&\left[\hat{\sigma}_j^{(2,0)}(0,v)-\tilde{\sigma}_j^{(2,0)}(0,v)\right]\\
&=T_j^{-1}\sum_{t=j+1}^{T}(\hat{z}_t-z_t)^2\hat{\psi}_{t-j}(v)+2T_j^{-1}\sum_{t=j+1}^{T}(\hat{z}_t-z_t)z_t\hat{\psi}_{t-j}(v)+T_j^{-1}\sum_{t=j+1}^{T}(z_t^2-1)[\hat{\psi}_{t-j}(v)-\psi_{t-j}(v)]\\
&=\hat{B}_{1j}(v)+2\hat{B}_{2j}(v)+\hat{B}_{3j}(v),\text{say.}
\end{aligned} \tag{A.10}$$

It follows from (A.10) that

$$\hat{A}_1\le 4\sum_{a=1}^{3}\sum_{j=1}^{T-1}k^2(j/p)T_j\int|\hat{B}_{aj}(v)|^2\,dW(v).$$

Proposition A.1 follows from Lemmas A.1–A.3 and $p/T\to 0$.

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

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

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

Proof of Proposition A.2: Given the decomposition in (A.10), we have

$$\left|\hat{\sigma}_j^{(2,0)}(0,v)-\tilde{\sigma}_j^{(2,0)}(0,v)]\tilde{\sigma}_j^{(2,0)}(0,v)^*\right| \le \sum_{a=1}^{3} |\hat{B}_{aj}(v)||\tilde{\sigma}_j^{(2,0)}(0,v)|,$$

where $\hat{B}_{aj}(v)$ are defined in (A.10). Therefore, Proposition A.2 follows from Lemmas A.4–A.6 and $p\to\infty, p/T\to 0$ as $T\to\infty$.

Lemma A.4: $\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{1j}(v)||\tilde{\sigma}_j^{(2,0)}(0,v)|\,dW(v) = O_P(p/T^{\frac{1}{2}}) = o_P(p^{\frac{1}{2}})$.

Lemma A.5: $\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{2j}(v)||\tilde{\sigma}_j^{(2,0)}(0,v)|\,dW(v) = O_P(1)+O_P(p/T^{\frac{1}{2}}) = o_P(p^{\frac{1}{2}})$.

Lemma A.6: $\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{3j}(v)||\tilde{\sigma}_j^{(2,0)}(0,v)|\,dW(v) = O_P(p/T^{\frac{1}{2}}) = o_P(p^{\frac{1}{2}})$.

Proof of Theorem A.2: The proof is similar to that of Theorem A.1. Let $\hat{A}_{1q}$ and $\hat{A}_{2q}$ be defined in the same way as $\hat{A}_1$ and $\hat{A}_2$ in (A.7), with $\{z_{q,t}\}_{t=1}^T$ replacing $\{\hat{z}_t\}_{t=1}^T$. It suffices to show $p^{-\frac{1}{2}}\hat{A}_{1q}\xrightarrow{p}0$ and $p^{-\frac{1}{2}}\hat{A}_{2q}\xrightarrow{p}0$. Put $\psi_{q,t}(v)\equiv e^{\mathbf{i}vz_{q,t}}-\phi_q(v)$, where $\phi_q(v)\equiv E(e^{\mathbf{i}vz_{q,t}})$. Let $\tilde{\sigma}_{q,j}^{(2,0)}(0,v)$ be defined as $\tilde{\sigma}_j^{(2,0)}(0,v)$, with $\{z_{q,t}\}_{t=1}^T$ replacing $\{z_t\}_{t=1}^T$. Then, similar to (A.10), we have

$$\begin{aligned}
&-\left[\tilde{\sigma}_j^{(2,0)}(0,v)-\tilde{\sigma}_{q,j}^{(2,0)}(0,v)\right]\\
&= T_j^{-1}\sum_{t=j+1}^{T}(z_t-z_{q,t})^2\psi_{q,t-j}(v)+2T_j^{-1}\sum_{t=j+1}^{T}(z_t-z_{q,t})z_{q,t}\psi_{q,t-j}(v)+T_j^{-1}\sum_{t=j+1}^{T}(z_t^2-1)[\psi_{q,t-j}(v)-\psi_{t-j}(v)]\\
&= \hat{B}_{1j,q}(v)+2\hat{B}_{2j,q}(v)+\hat{B}_{3j,q}(v),\text{say.} \qquad \text{(A.11)}
\end{aligned}$$

First, we consider $\hat{B}_{1j,q}(v)$. From the Cauchy-Schwarz inequality, we have

$$|\hat{B}_{1j,q}(v)|^2 \le \left[T_j^{-1}\sum_{t=j+1}^{T}(z_t-z_{q,t})^4\right]\left[T_j^{-1}\sum_{t=j+1}^{T}\left|\psi_{q,t-j}(v)\right|^2\right],$$

and from Markov's inequality and Assumption A.8, we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{1j,q}(v)|^2\,dW(v) \le \sum_{j=1}^{T-1} k^2(j/p)T_j\left[T_j^{-1}\sum_{t=j+1}^{T}(z_t-z_{q,t})^2\right]^2\int dW(v)\\
&= O_P(Tp/q^{2\kappa}). \qquad \text{(A.12)}
\end{aligned}$$

Next, we consider $\hat{B}_{2j,q}(v)$:

$$\left|\hat{B}_{2j,q}(v)\right|^2 \le \left[T_j^{-1}\sum_{t=j+1}^{T}(z_t-z_{q,t})^2\right]\left(T_j^{-1}\sum_{t=j+1}^{T}z_{q,t}^2\right)=O_P(q^{-\kappa}).$$

Thus, we have

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{2j,q}(v)|^2\,dW(v) \le \sum_{j=1}^{T-1} k^2(j/p)T_j\left[T_j^{-1}\sum_{t=j+1}^{T}(z_t-z_{q,t})^2\right]\left(T_j^{-1}\sum_{t=j+1}^{T}z_{q,t}^2\right)\int dW(v)\\
&= O_P(Tp/q^{\kappa}). \qquad \text{(A.13)}
\end{aligned}$$

Finally, we consider $\hat{B}_{3j,q}(v)$. According to the *m.d.s.* property of $\{z_t^2-1\}$ and the inequality $|e^{\mathbf{i}z_1}-e^{\mathbf{i}z_2}|\le|z_1-z_2|$ for any real z_1 and z_2, we have

$$E\left|\hat{B}_{3j,q}(v)\right|^2 \le T_j^{-2}\sum_{t=j+1}^{T}E\left(z_t^2-1\right)^2\left|\psi_{q,t-j}(v)-\psi_{t-j}(v)\right|^2$$

$$\le T_j^{-2}\sum_{t=j+1}^{T} E\left(z_t^2-1\right)^2\left|\psi_{q,t-j}(v)-\psi_{t-j}(v)\right|^2$$
$$\le v^2 T_j^{-1} E\left(z_{q,t-j}-z_{t-j}\right)^2.$$

Thus, from Assumptions A.7 and A.8, we have

$$\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{3j,q}(v)|^2\, dW(v) \le \sum_{j=1}^{T-1} k^2(j/p)\left[T_j^{-1}\sum_{t=j+1}^{T} E\left(z_{q,t-j}-z_{t-j}\right)^2\right]\int v^2 dW(v)$$
$$= O_P(p/q^{\kappa}). \tag{A.14}$$

By combining (A.11)–(A.14), we obtain

$$\hat{A}_{1q} \le \sum_{a=1}^{3}\sum_{j=1}^{T-1} k^2(j/p)T_j\int\left|\hat{B}_{aj,q}(v)\right|^2 dW(v) = O_P(Tp/q^{\kappa}).$$

Thus, given Assumption A.8, $q = p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$, we can obtain $p^{-\frac{1}{2}}\hat{A}_{1q} \xrightarrow{p} 0$. Moreover, from the Cauchy-Schwarz inequality, we can obtain

$$\sum_{j=1}^{T-1} k^2(j/p)T_j \int\left|\hat{B}_{aj,q}(v)\right|\left|\tilde{\sigma}_{qj}^{(2,0)}(0,v)\right| dW(v)$$
$$\le \left[\sum_{j=1}^{T-1} k^2(j/p)T_j\int\left|\hat{B}_{aj,q}(v)\right|^2 dW(v)\right]^{1/2}\left[\sum_{j=1}^{T-1} k^2(j/p)T_j\int\left|\tilde{\sigma}_{qj}^{(2,0)}(0,v)\right|^2 dW(v)\right]^{1/2}$$
$$= O_P(T^{\frac{1}{2}}p^{\frac{1}{2}}/q^{\frac{1}{2}\kappa})O_P(p^{\frac{1}{2}}) = O_P(T^{\frac{1}{2}}p/q^{\frac{1}{2}\kappa}).$$

Thus,

$$p^{-\frac{1}{2}}\hat{A}_{2q} = p^{-\frac{1}{2}}\sum_{a=1}^{3}\sum_{j=1}^{T-1} k^2(j/p)T_j Re\int \hat{B}_{aj,q}(v)\tilde{\sigma}_{qj}^{(2,0)}(0,v)^* dW(v)$$
$$= O_P(T^{\frac{1}{2}}p^{\frac{1}{2}}/q^{\frac{1}{2}\kappa}) = o_P(1),$$

given $q = p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$. This completes the proof of Theorem A.2.

Proof of Theorem A.3: The proof follows closely the proof of Theorem A.3 of Hong and Lee (2005) with $\{\varepsilon_{q,t},\psi_{q,t}(v)\}$ replaced by $\{z_{q,t}^2-1, e^{\mathbf{i}vz_{q,t-j}}-Ee^{\mathbf{i}vz_{q,t-j}}\}$.

Proof of Theorem 2: We consider $\hat{M}(p)$ only. The proof of Theorem 2 consists of the proofs of Theorems A.4 and A.5.

Theorem A.4: *Under the conditions of Theorem 2,* $(p^{\frac{1}{2}}/T)[\hat{M}(p)-M(p)]\xrightarrow{p} 0$.

Theorem A.5: *Under the conditions of Theorem 2,*

$$(p^{\frac{1}{2}}/T)M(p)\xrightarrow{p}\left[2D\int_0^{\infty}k^4(z)dz\right]^{-\frac{1}{2}}\int_{-\pi}^{\pi}|f^{(0,2,0)}(\omega,0,v)-f_0^{(0,2,0)}(\omega,0,v)|^2\, d\omega dW(v).$$

Proof of Theorem A.4: It suffices to show that

$$T^{-1}\int\sum_{j=1}^{T-1} k^2(j/p)T_j\left[|\hat{\sigma}_j^{(2,0)}(0,v)|^2-|\tilde{\sigma}_j^{(2,0)}(0,v)|^2\right]dW(v)\xrightarrow{p} 0, \tag{A.15}$$

$p^{-1}[\hat{C}(p)-\tilde{C}(p)] = O_P(1)$, and $p^{-1}[\hat{D}(p)-\tilde{D}(p)]\xrightarrow{p} 0$, where $\tilde{C}(p)$ and $\tilde{D}(p)$ are defined in the same way as $\hat{C}(p)$ and $\hat{D}(p)$ in (3.14), with $\{z_t\}_{t=1}^T$ replacing $\{\hat{z}_t\}_{t=1}^T$. Since the proofs of $p^{-1}[\hat{C}(p)-\tilde{C}(p)]=O_P(1)$ and $p^{-1}[\hat{D}(p)-\tilde{D}(p)]\xrightarrow{p}0$ are straightforward, we focus on the proof of (A.15). From the Cauchy-Schwarz inequality and the fact that $T^{-1}\int\sum_{j=1}^{T-1}k^2(j/p)T_j\,|\tilde{\sigma}_j^{(2,0)}(0,v)|^2\,dW(v)=O_P(1)$ as is implied by Theorem A.5 (the proof of Theorem A.5 does not depend on Theorem A.4), it suffices to show that $T^{-1}\hat{A}_1\xrightarrow{p}0$, where $\hat{A}_1$ is defined as in (A.8). Given (A.10), we show that $T^{-1}\int\sum_{j=1}^{T-1}k^2(j/p)T_j\,|\hat{B}_{aj}(v)|^2\,dW(v)\xrightarrow{p}0,\quad a=1,2,3.$

We first consider $a=1$. From the Cauchy-Schwarz inequality and $|\hat{\psi}_t(v)|\le 2$, we have

$$|\hat{B}_{1j}(v)|^2\le 2\left[T_j^{-1}\sum_{t=j+1}^{T}(\hat{z}_t-z_t)^2\right]^2.$$

It follows from (A.5) that

$$T^{-1}\int\sum_{j=1}^{T-1}k^2(j/p)T_j\,|\hat{B}_{1j}(v)|^2\,dW(v)\le\left[T_j^{-1}\sum_{t=1}^{T}(\hat{z}_t-z_t)^2\right]^2\sum_{j=1}^{T-1}k^2(j/p)\int dW(v)=O_P(p/T^2).$$

Next, we consider $a=2$. From the Cauchy-Schwarz inequality and $|\hat{\psi}_t(v)|\le 2$, we have

$$|\hat{B}_{2j}(v)|^2\le 2\left[T_j^{-1}\sum_{t=j+1}^{T}(\hat{z}_t-z_t)^2\right]\left[T_j^{-1}\sum_{t=j+1}^{T}z_t{}^2\right]=O_P(T^{-1}).$$

Finally, we consider $a=3$. We decompose it first as

$$\begin{aligned}\hat{B}_{3j}(v)&=T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)[\hat{\psi}_{t-j}(v)-\psi_{t-j}(v)]\\&=T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)[e^{\mathbf{i}v\hat{z}_{t-j}}-e^{\mathbf{i}vz_{t-j}}]-T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)[\hat{\phi}(v)-\phi(v)]\\&\equiv\hat{B}_{31j}(v)+\hat{B}_{32j}(v).\end{aligned}\tag{A.16}$$

From the Cauchy-Schwarz inequality and the inequality $|e^{\mathbf{i}z_1}-e^{\mathbf{i}z_2}|\le|z_1-z_2|$ for any real z_1 and z_2,

$$\begin{aligned}\left|\hat{B}_{31j}(v)\right|&=\left|T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)[e^{\mathbf{i}v\hat{z}_{t-j}}-e^{\mathbf{i}vz_{t-j}}]\right|\\&\le T_j^{-1}\sum_{t=j+1}^{T}\left|z_t^2-1\right||v|\left|\hat{z}_{t-j}-z_{t-j}\right|\\&\le|v|\left[T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)^2\right]^{1/2}\left[T_j^{-1}\sum_{t=j+1}^{T}\left(\hat{z}_{t-j}-z_{t-j}\right)^2\right]^{1/2}\\&=O_P(1)O_P(T^{-1/2})=O_P(T^{-1/2}),\end{aligned}\tag{A.17}$$

and by using the fact that $E\left|T^{-1}\sum_{t=1}^{T}e^{\mathbf{i}v\hat{z}_t}-\phi(v)\right|^2\le CT^{-1}$,

$$\left|\hat{B}_{32j}(v)\right| = \left|T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)\left[\hat{\phi}(v)-\phi(v)\right]\right|$$

$$\leq \left|\hat{\phi}(v)-\phi(v)\right|\left|T_j^{-1}\sum_{t=j+1}^{T}\left(z_t^2-1\right)\right|$$

$$= O_P(T^{-1/2})O_P(T^{-1/2}) = O_P(T^{-1}). \tag{A.18}$$

Combining (A.16)–(A.18) and $\sum_{j=1}^{T-1}k^2(j/p) = O(p)$ then yields the result of Theorem A.4.

Proof of Theorem A.5: The proof is similar to Hong (1999, Proof of Theorem 5), for the case $(m,l)=(2,0)$ and $W(\cdot)=\delta(\cdot)$, the Dirac delta function.

Proof of Theorem 3: Again, we only consider $\hat{M}(p)$. We show Theorems A.6 and A.7.

Theorem A.6: *Under the conditions of Theorem 3,* $\hat{M}(\hat{p})-M(\hat{p})\xrightarrow{p}0$.

Theorem A.7: *Under the conditions of Theorem 3,* $M(\hat{p})-M(p)\xrightarrow{p}0$.

Proof of Theorem A.6: Put $\hat{B}\equiv\sum_{j=1}^{T-1}k^2(j/\hat{p})T_j\int[|\hat{\sigma}_j^{(2,0)}(0,v)|^2-|\tilde{\sigma}_j^{(2,0)}(0,v)|^2]dW(v)$. It suffices to show $p^{-\frac{1}{2}}\hat{B}\xrightarrow{p}0$, $p^{-\frac{1}{2}}[\hat{C}_1(\hat{p})-\tilde{C}_1(\hat{p})]\xrightarrow{p}0$, and $p^{-1}[\hat{D}(\hat{p})-\tilde{D}(\hat{p})]\xrightarrow{p}0$. We show $p^{-\frac{1}{2}}\hat{B}\xrightarrow{p}0$; the proofs of $p^{-\frac{1}{2}}[\hat{C}(\hat{p})-\tilde{C}(\hat{p})]\xrightarrow{p}0$ and $p^{-1}[\hat{D}(\hat{p})-\tilde{D}(\hat{p})]\xrightarrow{p}0$ are similar. Given the conditions on $k(\cdot)$, there exists a symmetric monotonic decreasing function $k_0(z)$ in $z>0$ such that $|k(z)| \leq k_0(z)$ for all $z>0$, and $k_0(\cdot)$ satisfies Assumption A.5. It follows that for any constant $\varepsilon,\eta>0$,

$$P\left(p^{-\frac{1}{2}}|\hat{B}|>\varepsilon\right)\leq P\left(p^{-\frac{1}{2}}|\hat{B}|>\varepsilon,|\hat{p}/p-1|\leq\eta\right)+P\left(|\hat{p}/p-1|>\eta\right),$$

where the second term vanishes for all $\eta>0$ asymptotically given $\hat{p}/p-1\xrightarrow{p}0$. Thus, it remains to show that the first term also vanishes as $T\to\infty$.

Because $|\hat{p}/p-1|\leq\eta$ implies $\hat{p}\leq(1+\eta)p$, we have that for $|\hat{p}/p-1|\leq\eta$,

$$p^{-\frac{1}{2}}|\hat{B}|\leq(1+\eta)^{\frac{1}{2}}[(1+\eta)p]^{-\frac{1}{2}}\sum_{j=1}^{T-1}k_0^2[j/(1+\eta)p]T_j\left[|\hat{\sigma}_j^{(2,0)}(0,v)|^2-|\tilde{\sigma}_j^{(2,0)}(0,v)|^2\right]\xrightarrow{p}0$$

for any $\eta>0$ given (A.10), where the inequality follows from the fact that $|k(z)|\leq|k_0(z)|$. This completes the proof of Theorem A.6.

Proof of Theorem A.7: Put $\tilde{Q}(p)\equiv 2\pi\sum_{j=1}^{T-1}k^2(j/\hat{p})T_j\int[|\tilde{\sigma}_j^{(2,0)}(0,v)|^2\,dW(v)$. Then, we can write

$$M(\hat{p})-M(p)=\left[\tilde{Q}(\hat{p})-\tilde{C}(\hat{p})\right]/\sqrt{\tilde{D}(\hat{p})}-\left[\tilde{Q}(p)-\tilde{C}(p)\right]/\sqrt{\tilde{D}(p)}$$

$$=\left\{[\tilde{Q}(\hat{p})-\tilde{Q}(p)]-[\tilde{C}(\hat{p})-\tilde{C}(p)]\right\}/\sqrt{\tilde{D}(\hat{p})}+M(p)\left[\sqrt{\tilde{D}(p)/\tilde{D}(\hat{p})}-1\right].$$

Following a reasoning analogous to the proof of Theorem 4 in Hong (1999), we can obtain $p^{-\frac{1}{2}}[\tilde{Q}(\hat{p})-\tilde{Q}(p)]\xrightarrow{p}0$ under H_0. This and Lemma A.7 imply $M(\hat{p})-M(p)\xrightarrow{p}0$. Hence, $M(\hat{p})\xrightarrow{d}N(0,1)$.

Lemma A.7: Suppose Assumptions A.5 and A.6 hold. If $\hat{p}/p=1+O_P(p^{-(\frac{3\beta}{2}-1)})$ for some $\beta>(2b-\frac{1}{2})/(2b-1)$, where b is as in Assumption A.6 and $p=cT^{\lambda}$ for $\lambda\in(0,\frac{2b-1}{4b-1})$ and $c\in(0,\infty)$. Then, $p^{-\frac{1}{2}}[\hat{C}(\hat{p})-\tilde{C}(p)]\xrightarrow{p}0$ and $p^{-1}[\hat{D}(\hat{p})-\tilde{D}(p)]\xrightarrow{p}0$.

References

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3–30.

Bera, A.K., Higgins, M.L., 1993. ARCH models: Properties, estimations and testing. Journal of Economic Surveys 7, 305–362.

Berkes, I., Horváth, L., Kokoszka, P., 2003. Asymptotics for GARCH squared residual correlations. Econometric Theory 19, 515–540.

Bierens, H., 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Chou, T.R., Kroner, K., 1992. ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5–59.

Bollerslev, T., Engle, R.F., 1988. A capital asset pricing model with time-varying covariances. Journal of Political Economy 96, 116–131.

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models, in: Engle, R.F, McFadden, D. (Eds), Handbook of Econometrics. North Holland Press, Amsterdam, the Netherlands, pp. 2959–3038.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. Econometric Reviews 11, 143–172.

Box, G.E.P., Pierce, D.A., 1970. Distribution of the residual autocorrelations in ARIMA time series models. Journal of the American Statistical Association 65, 1509–1526.

Brillinger, D.R., 1981. Time Series: Data Analysis and Theory. Holden-Day, San-Fransisco, CA, USA.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

Drost, F.C., Nijman, T.E., 1993. Temporal aggregation of GARCH processes. Econometrica 61, 909–927.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. The Journal of Finance 47, 577–605.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the U.K. Inflation. Econometrica 50, 987–1008.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 5, 1749–1778.

Gallant, A.R., Hsieh, D., Tauchen, G., 1991. On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-83, in: Barnett, W.A., Powell, J., Tauchen, G.E. (Eds), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, pp. 199–240.

Gallant, A.R., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. Econometrica 57, 1091–1120.

Geweke, J., 1994. Comment on Bayesian analysis of stochastic volatility. Journal of Business and Economic Statistics 12, 371–417.

Glosten, L.R., Jaganathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Hafner, C., 1998. Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility. Heidelberg, New York, Physica-Verlag.

Hamilton, J.D., Lin, G., 1996. Stock market volatility and the business cycle. Journal of Applied Econometrics 11, 573–593.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

Harvey, C., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34, 465–487.

Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1296.

Higgins, M.L., Bera, A.K., 1992. A class of nonlinear ARCH models. International Economic Review 33, 137–158.

Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. Journal of American Statistical Association 84, 1201–1220.

Hong, Y., Lee, Y.J., 2005. Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. Reviews of Economic Studies 72, 499–541.

Horváth, L., Kokoszka, P., 2001. Large sample distribution of weighted sums of ARCH(p) squared residual correlations. Econometric Theory 17, 283–295.

Hsiao, C., Li, Q., 2001. A consistent test for conditional heteroskedasticity in time-series regression models. Econometric Theory 17, 188–221.

Jacquier, E., Polson, N., Rossi, P., 2004. Bayesian analysis of stochastic volatility models with leverage effect and fat tails. Journal of Econometrics 122, 185–212.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 27, 1699–1737.

Li, W.K., Mak, T., 1994. On the squared autocorrelations in non-linear time series with conditional heteroskedasticity. Journal of Time Series Analysis 15, 627–636.

Lumsdaine, R.L., 1996. Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. Econometrica 64, 575–596.

Lundbergh, S., Teräsvirta, T., 2002. Evaluating GARCH models. Journal of Econometrics 110, 417–435.

Mikosch, T., Starica, C., 2000. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Annals of Statistics 28, 1427–1451.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–370.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press, London, UK.

Sentana, E., 1995. Quadratic ARCH models. Reviews of Economic Studies 62, 639–661.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London, UK.

Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295–324.

Tsay, R.S., 2002. Analysis of Financial Time Series. John Wiley and Sons, Inc., New York, NY, USA.

Tse, Y.K., 2002. Residual-based diagnostics for conditional heteroskedasticity models. Econometrics Journal 5, 358–373.

Weiss, A.A., 1986. Asymptotic theory for ARCH models: Estimation and testing. Econometric Theory 2, 107–131.

White, H., 1999. Asymptotic Theory for Econometricians (2nd ed.). Academic Press, San Diego, CA, USA.

Wooldridge, J., 1990. A unified approach to robust, regression-based specification tests. Econometric Theory 6, 17–43.

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.

Zakoian, J.M., 1994. Threshold heteroskedasticity models. Journal of Economics Dynamics and Control 15, 931–955.