



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

The Review of Economic Studies, Ltd.

Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form

Author(s): Yongmiao Hong and Yoon-Jin Lee

Source: *The Review of Economic Studies*, Vol. 72, No. 2 (Apr., 2005), pp. 499-541

Published by: [Oxford University Press](#)

Stable URL: <http://www.jstor.org/stable/3700661>

Accessed: 22/11/2013 14:04

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Oxford University Press and *The Review of Economic Studies, Ltd.* are collaborating with JSTOR to digitize, preserve and extend access to *The Review of Economic Studies*.

<http://www.jstor.org>

Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form

YONGMIAO HONG

Cornell University and Tsinghua University

and

YOON-JIN LEE

Cornell University

First version received January 2003; final version accepted March 2004 (Eds.)

Economic theories in time series contexts usually have implications on and only on the conditional mean dynamics of underlying economic variables. We propose a new class of specification tests for time series conditional mean models, where the dimension of the conditioning information set may be infinite. Both linear and nonlinear conditional mean specifications are covered. The tests can detect a wide range of model misspecifications in mean while being robust to conditional heteroscedasticity and higher order time-varying moments of unknown form. They check a large number of lags, but naturally discount higher order lags, which is consistent with the stylized fact that economic behaviours are more affected by the recent past events than by the remote past events. No specific estimation method is required, and the tests have the appealing “nuisance parameter free” property that parameter estimation uncertainty has no impact on the limit distribution of the tests. A simulation study shows that it is important to take into account the impact of conditional heteroscedasticity; failure to do so will cause overrejection of a correct conditional mean model. In a horse race competition on testing linearity in mean, our tests have omnibus and robust power against a variety of alternatives relative to some existing tests. In an application, we find that after removing significant but possibly spurious autocorrelations due to nonsynchronous trading, there still exists significant predictable nonlinearity in mean for S&P 500 and NASDAQ daily returns.

1. INTRODUCTION

Nonlinear time series analysis has been advancing rapidly, with wide applications in economics and finance (*e.g.* Brock, Hsieh and LeBaron (1991), Granger and Teräsvirta (1993), Teräsvirta, Tjøstheim and Granger (1994b), Tjøstheim (1994), Granger (2001)). Like linear time series analysis, nonlinear time series modelling involves model identification, estimation and evaluation. Specification analysis is needed in model identification and evaluation. In this paper, we shall develop a new class of specification tests for conditional mean models in time series with conditional heteroscedasticity of unknown form.

Most economic theories in dynamic contexts, such as efficient market hypothesis, expectations hypothesis, consumption smoothing, dynamic asset pricing and rational expectations, have implications on and only on the conditional mean dynamics of underlying economic variables given the information available to economic agents (*e.g.* Cochrane (2001),

Sargent and Ljungqvist (2002)). For example, the conventional efficient market hypothesis states that the expected asset return given the information available, is zero, or at most, is constant over time (e.g. Fama (1970, 1991), Campbell, Lo and MacKinlay (1997, Chapter 1)), and dynamic asset pricing implies that the expectation of the pricing error given the information available is zero for all assets (Cochrane, 2001). Although economic theory may suggest a nonlinear relationship, it does not give a concrete functional form for the conditional mean of economic variables.¹ Various parametric models used in practice can be, at best, regarded as approximations to the underlying conditional mean dynamics. There are many ways to specify which variables in the information set, which functional forms, and which lag structures to be used in conditional mean modelling. It is important to test conditional mean specification, because misspecification in mean can lead to misleading conclusions on economic theories and hypotheses, and to suboptimal point forecasts.

In time series modelling, it is important to determine first whether a time series is linear in mean; *i.e.* whether the conditional mean of a process is a linear combination of the variables in an information set. There have been a variety of tests for linearity in mean in the literature. They can be divided into two broad categories. One contains the tests derived without a specific nonlinear alternative. Examples include Keenan (1985), Tsay (1986) and White (1989). The other category consists of tests against a specific nonlinear alternative, usually formulated as Lagrangian Multiplier (LM) or LM-type tests. Examples include Luukkonen, Saikkonen and Teräsvirta (1988*a,b*), Saikkonen and Luukkonen (1988), Teräsvirta, Lin and Granger (1994*a*), Hamilton (2001), and Dahl and Gonzalez-Rivera (2003). The LM tests of Hamilton (2001) and Dahl and Gonzalez-Rivera (2003) are based on the random field theory. These tests contain a rich class of alternative models and can be interpreted as tests based on spline smoothers (Dahl, 2002). The greatest appeal of the tests in the second group is that they involve no estimation of the specified nonlinear alternative model, which could otherwise be computationally intensive. On the other hand, some tests in the first group can be interpreted as LM tests against a nonlinear alternative. Such interpretation is informative because it can reveal against which type of nonlinearity a test has the best power. Of course, some tests with unspecified alternatives do not have an LM test interpretation, although they have the best power against certain implicit alternatives.

When there is prior information about a potential alternative, or when one is interested in certain specific alternatives, tests against a specific alternative are a natural choice. This is exactly the case for the tests by Luukkonen *et al.* (1988*a,b*) and Saikkonen and Luukkonen (1988), where the interest is in modelling smooth regime changes. In practice, economic theory often does not point to a single nonlinear alternative. In such scenarios, it is desirable to have an omnibus test against general departure from linearity. Except White's (1989) neural network test, which can detect all possible misspecified functional forms given a correct lag specification, other existing tests for linearity in mean are of parametric nature in the sense that they all test, explicitly or implicitly, a parametric alternative with a fixed lag specification. These tests have optimal power against a specific alternative but they may have low or little power against other alternatives.

In this paper, we propose a class of generally applicable omnibus tests for time series conditional mean models, with no prior knowledge of possible alternatives (including both functional forms and lag structures). Both linear and nonlinear conditional mean models are covered in a unified set-up. We use the generalized spectrum, which was proposed in

1. In some circumstances, economic theory may suggest certain restrictions on the functional form and/or the lag structure of the conditional mean dynamics of an underlying process. For example, Hall's (1978) consumption smoothing hypothesis suggests that the expectation of the next period's marginal utility of consumption given the information available at the current period is a function of the current period consumption only. However, the functional form is not specified unless the utility function of the representative economic agent is known.

Hong (1999) as a new analytic tool for nonlinear time series. Thanks to the use of the characteristic function, the generalized spectrum can capture both linear and nonlinear serial dependence.² The latter can be subtle and difficult to detect using conventional econometric tools. In the meantime, the generalized spectrum enjoys the nice features of spectral analysis. In particular, it incorporates information on the serial dependence from all lags and can characterize cyclical dynamics caused by linear or nonlinear serial dependence. Thus, our approach can detect a wide variety of misspecifications in both functional form and lag structure. This distinguishes our tests from the existing tests for time series conditional mean models, all of which assume a fixed lag order specification and focus on the functional form misspecification. One important feature of time series conditional mean modelling is that the conditioning information set usually contains an infinite number of lags (*i.e.* the entire past history), unless a Markovian assumption holds. Our tests check a large number of lags without suffering from the curse of dimensionality. When a large number of lags is used, chi-square tests for linearity usually have poor power in finite samples, due to the loss of a large number of degrees of freedom (see, *e.g.* Dahl and Gonzalez-Rivera, 2003, for more discussion). This undesired feature, fortunately, is not shared by our generalized spectral approach, because it naturally discounts higher order lags, which is consistent with the stylized fact that economic behaviours are usually more influenced by the recent events than by the remote past events. Thus, our tests are particularly useful when the information set has a large dimension. We note that our tests can be used to test the martingale hypothesis for observed raw data without any modification.

It should be emphasized that Hong's (1999) generalized spectrum itself cannot be used to test conditional mean specification, because it can capture the serial dependence in every conditional moment and thus cannot separate the serial dependence in mean from that in higher order moments. However, because the characteristic function can be differentiated to generate moments, a suitable generalized spectral *derivative* enables us to focus only on the serial dependence in mean, making it suitable to test conditional mean specification. We compare a nonparametric (inefficient) generalized spectral derivative estimator with a restricted (efficient) counterpart implied by correct conditional mean specification. Our tests can be viewed as a generalization of Hausman's (1978) method from a parametric context to a nonparametric time series context, although the asymptotic theory is different.

Most existing tests for time series conditional mean models assume observable explanatory variables. In contrast, our tests can check various linear and nonlinear parametric conditional mean models with possibly unobservable variables (*e.g.* past shocks) and an infinite number of lags. Examples include ARMA models, ARMAX models, regime-switching models (Hamilton, 1989), state-space models (Priestley, 1988), smooth transition autoregressive models (Teräsvirta, 1994), Poisson jump models, and threshold autoregressive models with known thresholds (Tiao and Tsay (1994), Potter (1995)). Our procedures only require as inputs estimated model residuals, obtained from any \sqrt{T} -consistent parameter estimates. We note that compared to the vast literature on testing linearity in mean, there are relatively few tests for nonlinear time series conditional mean models. Exceptions are Tsay (1989), Eitheim and Teräsvirta (1996), Hamilton (2001), Gao and King (2001), and Chen, Härdle and Li (2003).

Economic theory, while having implications on the conditional mean dynamics of an underlying process, is usually silent about its higher order conditional moment dynamics. Thus, as emphasized by Granger (1995), it is important to develop tests of conditional mean models that are robust to conditional heteroscedasticity and other higher order time-varying moments. Failure to accommodate conditional heteroscedasticity will lead to improper levels for the tests, giving

2. There has been an increasing interest in using the characteristic function in econometrics (*e.g.* Epps (1987, 1988), Pinkse (1998), Singleton (2001), Jiang and Knight (2002), Knight and Yu (2002), and Chacko and Viceira (2003)).

a misleading conclusion. For example, an ARCH process is similar to a bilinear autoregressive process in terms of autocorrelations in level and level-square, respectively (Bera and Higgins, 1997). A LM test for a bilinear alternative will be likely to mistake an ARCH process as a bilinear process if conditional homoscedasticity is assumed. As is well known (e.g. Diebold and Nason (1990), Meese and Rose (1991), Granger (1992, Section 8)), the distinction between nonlinearities in mean and in higher order moments has important economic implications. For example, suppose an asset return follows a bilinear process. Then the level of asset return is predictable using its past history. In contrast, if the asset return follows an ARCH process, then its level is not predictable because it is a martingale difference sequence (*m.d.s.*). As an important feature, our tests for conditional mean models are robust to conditional heteroscedasticity and all other higher order conditional moments of unknown form. Most existing tests for linearity in mean assume conditional homoscedasticity or i.i.d. errors (e.g. Keenan (1985), Tsay (1986), White (1989), Hamilton (2001), Dahl and Gonzalez-Rivera (2003)).

Section 2 introduces the hypotheses of interest and discusses the concept of (non)linearity in mean. Section 3 describes our approach and introduces our test statistics. Section 4 derives the asymptotic distribution of the tests, and Section 5 investigates their asymptotic power. Section 6 justifies the use of a data-driven lag order and considers a plug-in method. In Section 7, we examine the finite sample performance of the tests. In Section 8, we apply the tests to S&P 500 and NASDAQ daily indices. Section 9 concludes, and all proofs are collected in the Appendix. The GAUSS code for implementing our tests is available from the authors upon request. Throughout, we use C to denote a generic bounded constant, $\|\cdot\|$ the Euclidean norm, and A^* the complex conjugate of A .

2. FRAMEWORK AND NONLINEARITY IN MEAN

Suppose $\{Y_t\}$ is a strictly stationary process. We consider a time series model

$$Y_t = g(I_{t-1}, \theta) + \varepsilon_t, \quad t = 1, 2, \dots, \quad (2.1)$$

where I_{t-1} is an information set at time $t - 1$, which may contain lagged dependent variables $\{Y_{t-j}, j > 0\}$, lagged shocks $\{\varepsilon_{t-j}, j > 0\}$, and current and lagged exogenous variables $\{Z_{t-j}, j \geq 0\}$; $g(I_{t-1}, \theta)$ is a parametric model for $E(Y_t | I_{t-1})$, the conditional mean of Y_t given I_{t-1} , and $\theta \in \Theta$ is a finite-dimensional parameter. In time series modelling, I_{t-1} is possibly infinite-dimensional (*i.e.* dating back to the infinite past), as is the case for non-Markovian processes. This poses a challenge in testing adequacy of the model $g(I_{t-1}, \theta)$, due to the curse of dimensionality.

We say that the model $g(I_{t-1}, \theta)$ is correctly specified for $E(Y_t | I_{t-1})$ if there is some $\theta_0 \in \Theta$ such that

$$\mathbb{H}_0 : \Pr[g(I_{t-1}, \theta_0) = E(Y_t | I_{t-1})] = 1. \quad (2.2)$$

Alternatively, the model $g(I_{t-1}, \theta)$ is misspecified for $E(Y_t | I_{t-1})$ if for all $\theta \in \Theta$, we have

$$\mathbb{H}_A : \Pr[g(I_{t-1}, \theta) = E(Y_t | I_{t-1})] < 1. \quad (2.3)$$

Conditional mean modelling has been the primary interest in time series analysis, because $E(Y_t | I_{t-1})$ is the optimal predictor for Y_t using I_{t-1} in terms of the mean squared error criterion. In addition, as discussed earlier, most economic theories have implications on and only on the conditional mean dynamics of the underlying economic variable.

In time series analysis, testing linearity in mean has attracted much attention, which is often the first stage in nonlinear time series modelling (Granger and Teräsvirta (1993, Chapter 6),

Hansen (1999)). Suppose

$$g(I_{t-1}, \theta) = X_t' \Lambda(\theta), \quad (2.4)$$

where X_t is a $d \times 1$ vector in I_{t-1} , $\Lambda(\theta)$ is a $d \times 1$ parameter that depends on θ , and integer d can be finite or infinite. Following Granger and Teräsvirta (1993) and Lee, White and Granger (1993), we say that Y_t is linear in mean on X_t if there exists some $\theta_0 \in \Theta$ such that

$$\Pr[E(Y_t | I_{t-1}) = X_t' \Lambda(\theta_0)] = 1. \quad (2.5)$$

Otherwise, we say that model (2.4) suffers from neglected nonlinearity in mean. A more restrictive version of this definition was given in Granger and Teräsvirta (1993) and Lee *et al.* (1993), who consider the case that X_t is observable with a fixed dimension. Here, we allow X_t to be unobservable, possibly with an infinite dimension. This generalization is useful because, when X_t has a fixed dimension, the violation of (2.5) can be caused by neglected higher order lags rather than neglected nonlinearity. In other words, the alternative to (2.4) can be a linear time series with a higher order lag structure.

Our definition of linearity in mean differs from the usual notion of a linear time series in the literature, which is a weighted sum of current and past shocks $\{\varepsilon_{t-j}, j \geq 0\}$, where $\{\varepsilon_t\}$ is serially uncorrelated (*e.g.* Priestley, 1981, p. 141). Such a process may not be linear in mean in the sense of (2.5), because a white noise $\{\varepsilon_t\}$ may not be a *m.d.s.* A white noise process with nonzero conditional mean is predictable in mean. Only when $\{\varepsilon_t\}$ is a *m.d.s.*, the notion of a linear time series coincides with our definition of linearity in mean. The latter concept is more useful for modelling conditional mean dynamics.

Likewise, our concept of nonlinearity in mean differs from the conventional notion of a nonlinear time series. The latter means any departure from a linear time series with independent and identically distributed (i.i.d.) errors $\{\varepsilon_t\}$. According to this definition, an AR model with ARCH errors is a nonlinear time series. It is linear in mean, however, in view of (2.5). As we have emphasized, it is important to distinguish nonlinearities in different moments, which have different economic implications.

It should be noted that some popular “linearity” tests in the time series literature are designed to check departures from i.i.d. for observed raw data or the model error $\{\varepsilon_t\}$ (see Barnett, Gallant, Hinich, Jungeilges, Kaplan and Jensen, 1997, for a survey). They can capture nonlinearities in mean and in higher order moments, and therefore are not suitable for testing linearity in mean. For example, as pointed out in Lee *et al.* (1993), the well-known BDS test (Brock *et al.* (1991), Brock, Dechert, Scheinkman and LeBaron (1996)) can reject a correct linear AR model with ARCH errors. Similarly, the tests by Subba Rao and Gabr (1980) and Hinich (1982) are also not suitable for testing nonlinearity in mean. These tests exploit the bispectral shape of a linear process with i.i.d. errors and can detect nonlinearity in mean and in higher order moments.³

Existing parametric tests for linearity in mean are powerful against some misspecifications in mean, but they all suffer from power loss in detecting certain nonlinear alternatives (*e.g.* Tong, 1990, Chapter 5). When there is no prior information about the alternative, it is desirable to have an omnibus test against a wide range of alternatives. Tsay (2001, p. 268) suggested combining several procedures via an augmented alternative. This will ensure power against a wider range of alternatives than an individual test, but the choice of the augmented alternative is somewhat arbitrary. Another approach is to estimate $E(Y_t | I_{t-1})$ nonparametrically and compare the nonparametric estimator with $g(I_{t-1}, \theta)$, or equivalently, to estimate $E(\varepsilon_t | I_{t-1})$ nonparametrically and check if it is zero (Li (1999), Gao and King (2001)). This approach works

3. Although the generalized spectrum shares certain features similar to the bispectrum, our generalized spectral derivative introduced below focuses on and only on serial dependence in mean. Thus, it is suitable for testing conditional mean specification.

when $E(Y_t | I_{t-1}) = E(Y_t | X_t)$ *a.s.* and the dimension of $X_t \in I_{t-1}$ is small. However, it is not expected to work well in finite samples when the dimension of X_t is large. Our generalized spectral approach provides a solution to this difficulty of “curse of dimensionality”. It checks many lags in a pairwise manner, which avoids the “curse of dimensionality”.⁴ This is similar in spirit to the nonparametric additive time series modelling strategy in the literature (*e.g.* Marsry and Tjøstheim (1997), Gao, Tong and Wolff (2002), Kim and Linton (2003)). On the other hand, the generalized spectral derivative can detect various misspecified functional forms at any given lag order. One can view our procedures, when applied to linearity testing, as checking the joint hypotheses of (i) $E(Y_t | I_{t-1}) = E(Y_t | X_t)$ for some $X_t \in I_{t-1}$ and (ii) $E(Y_t | X_t) = X_t' \theta_0$ for some θ_0 . All existing tests for linearity in mean only focus on testing (ii) and ignore testing (i); they can easily miss conditional mean misspecifications that occur at higher lag orders. Moreover, even for any given lag order, the existing linearity tests except White’s (1989) neural network test cannot detect all departures from (i).

3. APPROACH AND TEST STATISTICS

3.1. Generalized spectral analysis

Our approach to testing correct conditional mean specification (\mathbb{H}_0 in (2.2)) is based on Hong’s (1999) generalized spectrum, which is an analytic tool for nonlinear time series, just as the power spectrum is an analytic tool for linear time series (Priestley, 1981).

Recall the model error

$$\varepsilon_t(\theta) \equiv Y_t - g(I_{t-1}, \theta) \quad (3.1)$$

has the property that $E[\varepsilon_t(\theta_0) | I_{t-1}] = 0$ *a.s.* for some $\theta_0 \in \Theta$. This implies

$$E[\varepsilon_t(\theta_0) | I_{t-1}^\varepsilon] = 0 \quad \textit{a.s.}, \quad (3.2)$$

where $I_{t-1}^\varepsilon \equiv \{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0), \dots\}$. Thus, to test \mathbb{H}_0 , we can check if $E[\varepsilon_t(\theta_0) | I_{t-1}^\varepsilon] = 0$ *a.s.*⁵ Still, we have the curse of dimensionality problem because I_{t-1}^ε has an infinite dimension. Fortunately, the generalized spectral approach provides a sensible way to tackle this difficulty.

For notational economy, we put $\varepsilon_t \equiv \varepsilon_t(\theta)$. Suppose $\{\varepsilon_t\}$ is a strictly stationary process with marginal characteristic function $\varphi(u) \equiv E(e^{iu\varepsilon_t})$ and pairwise joint characteristic function $\varphi_j(u, v) \equiv E(e^{iu\varepsilon_t + iv\varepsilon_{t-j}})$, where $i \equiv \sqrt{-1}$, $u, v \in \mathbb{R}$, and $j = 0, \pm 1, \dots$. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{iu\varepsilon_t}\}$. It is defined as

$$f(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \quad (3.3)$$

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

$$\sigma_j(u, v) \equiv \text{cov}(e^{iu\varepsilon_t}, e^{iv\varepsilon_{t-j}}), \quad j = 0, \pm 1, \dots$$

The function $f(\omega, u, v)$ can capture any type of pairwise serial dependence in $\{\varepsilon_t\}$, *i.e.* dependence between ε_t and ε_{t-j} for any nonzero lag j , including that with zero autocorrelation. This is analogous to the higher order spectra (Brillinger and Rosenblatt, 1967*a,b*) in the sense that $f(\omega, u, v)$ can capture the serial dependence in higher order moments. However, unlike

4. From a theoretical point of view, the pairwise approach will miss dependent processes that are pairwise independent. However, such processes apparently do not appear in most empirical applications in economics and finance.

5. For a univariate time series, the knowledge of $I_{t-1} \equiv \{Y_{t-1}, Y_{t-2}, \dots\}$ is equivalent to the knowledge of $I_{t-1}^\varepsilon \equiv \{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0), \dots\}$ under certain regularity conditions (*e.g.* Priestley, 1988, p. 72). However, when I_{t-1} contains other current and lagged exogenous variables, these two information sets generally differ. Below, we will first develop tests based on I_{t-1}^ε , and then consider extensions to the more general information set in Section 5.

the higher order spectra, $f(\omega, u, v)$ does not require existence of any moment of $\{\varepsilon_t\}$. This is important in economics and finance because it has been argued that the higher order moments of many financial time series may not exist.

When $\sigma^2 \equiv E(\varepsilon_t^2)$ exists, we can obtain the power spectrum as a derivative of $f(\omega, u, v)$:

$$-\frac{\partial^2}{\partial u \partial v} f(\omega, u, v) |_{(u,v)=(0,0)} = h(\omega) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \text{cov}(\varepsilon_t, \varepsilon_{t-|j|}) e^{-ij\omega}, \quad \omega \in [-\pi, \pi].$$

For this reason, we call $f(\omega, u, v)$ the generalized spectrum of $\{\varepsilon_t\}$.

As is well known, the interpretation of spectral analysis is much more difficult for nonlinear time series than for linear time series. For example, the bispectrum has no physical (e.g. energy decomposition over frequencies) interpretation, unlike the power spectrum $h(\omega)$. This is also the case for the generalized spectrum $f(\omega, u, v)$. However, the basic idea of characterizing cyclical dynamics still applies: $f(\omega, u, v)$ has useful interpretations when searching for linear or nonlinear cycles. A strong cyclicity of data may be linked with a strong serial dependence in $\{\varepsilon_t\}$ that may not be captured by the autocorrelation function. The generalized spectrum $f(\omega, u, v)$ can capture such nonlinear cyclical patterns by displaying distinct spectral peaks. For example, suppose an asset return series has a stochastic cyclical dynamics in volatility, which may be linked to business cycles (e.g. Hamilton and Lin, 1996). Then the power spectrum $h(\omega)$ will be flat and miss the volatility cycles. In contrast, $f(\omega, u, v)$ can effectively capture such cycles. More generally, the generalized spectrum can capture cyclical dynamics caused by linear and nonlinear dependence. The latter includes the serial dependence in volatility, skewness and other higher order conditional moments.⁶

The generalized spectrum $f(\omega, u, v)$ itself is not suitable for testing \mathbb{H}_0 in (2.2), because it can capture the serial dependence in mean and in higher order moments. An example is an ARCH process. The generalized spectrum $f(\omega, u, v)$ can capture this process, although it is a *m.d.s.* However, just as the characteristic function can be differentiated to generate various moments of $\{\varepsilon_t\}$, $f(\omega, u, v)$ can be differentiated to capture the serial dependence in various moments. To capture (and only capture) the serial dependence in the conditional mean, one can use the derivative

$$f^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{(1,0)}(0, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \quad (3.4)$$

where

$$\sigma_j^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u} \sigma_j(u, v) |_{u=0} = \text{cov}(i\varepsilon_t, e^{iv\varepsilon_{t-|j|}}).$$

The measure $\sigma_j^{(1,0)}(0, v)$ checks whether the autoregression function $E(\varepsilon_t | \varepsilon_{t-j})$ at lag j is zero. Under appropriate conditions, $\sigma_j^{(1,0)}(0, v) = 0$ for all $v \in \mathbb{R}$ if and only if $E(\varepsilon_t | \varepsilon_{t-j}) = 0$ *a.s.*⁷ The autoregression function can capture linear and nonlinear serial dependence in mean, including the processes with zero autocorrelation. Examples are a bilinear autoregressive process $\varepsilon_t = \alpha z_{t-1} \varepsilon_{t-2} + z_t$ and a nonlinear moving-average process $\varepsilon_t = \alpha z_{t-1} z_{t-2} + z_t$, where $\{z_t\} \sim \text{i.i.d.}(0, \sigma^2)$. These processes are white noises, but they are not a *m.d.s.*, because their conditional means are time-varying. Thus, $E(\varepsilon_t | \varepsilon_{t-j})$ is a natural tool to test \mathbb{H}_0 , whereas $\text{cov}(\varepsilon_t, \varepsilon_{t-j})$ will miss such subtle nonlinear processes. Nevertheless, $E(\varepsilon_t | \varepsilon_{t-j})$ has not been

6. A potentially useful application is the investigation of possible nonlinear business cycles by $f(\omega, u, v)$. The power spectrum $h(\omega)$, when applied to macroeconomic time series such as the U.S. GDP growth rates, often produces a flat spectrum. However, some nonlinear time series experts (e.g. Tong, 1990, p. 232) believe that business cycles are related to nonlinear cyclical dynamics. It will be interesting to examine whether $f(\omega, u, v)$ can capture and identify such nonlinear cycles.

7. See Bierens (1982) and Stinchcombe and White (1998) for discussion in a different context with i.i.d. samples.

as widely used as $\text{cov}(\varepsilon_t, \varepsilon_{t-j})$. An exception is Hjellvik and Tjøstheim (1996), who considered testing linearity in mean for observed raw data using a kernel estimator for the autoregression function. Tong (1990) and Teräsvirta *et al.* (1994a) also discussed smoothed nonparametric estimation of the autoregression function.

Although $E(\varepsilon_t | \varepsilon_{t-j})$ and $\sigma_j^{(1,0)}(0, v)$ are equivalent measures, the use of $\sigma_j^{(1,0)}(0, v)$ avoids smoothed nonparametric estimation. In addition, the measure $\sup_{v \in \mathbb{R}} |\sigma_j^{(1,0)}(0, v)|$ can be viewed as an operational version of the maximum mean correlation, $\max_{f(\cdot)} |\text{corr}[\varepsilon_t, f(\varepsilon_{t-j})]|$, which was proposed by Granger and Teräsvirta (1993, p. 23) as a measure for nonlinearity in mean. Similarly, the supremum generalized spectral derivative modulus

$$m(\omega) \equiv \sup_{v \in (-\infty, \infty)} |f^{(0,1,0)}(\omega, 0, v)|, \quad \omega \in [-\pi, \pi], \tag{3.5}$$

can be viewed as the maximum dependence in mean at frequency ω . It can be used to search cycles in mean that are caused by linear or nonlinear serial dependence in mean. An example of the latter is the well-known ARCH-in-mean effect (Engle, Lilien and Robins, 1987).

The hypothesis of $E(\varepsilon_t | I_{t-1}^\varepsilon) = 0$ *a.s.* is not the same as the hypothesis of $E(\varepsilon_t | \varepsilon_{t-j}) = 0$ *a.s.* for all $j > 0$. The former implies the latter but not vice versa. This is the price we have to pay for dealing with the difficulty of the “curse of dimensionality”. One example that is not a *m.d.s.* but has $E(\varepsilon_t | \varepsilon_{t-j}) = 0$ *a.s.* for all $j > 0$ is a nonlinear moving-average process⁸

$$\varepsilon_t = \alpha z_{t-2} z_{t-3} + z_t, \quad \{z_t\} \sim \text{i.i.d. } (0, \sigma^2). \tag{3.6}$$

Obviously, there are many such examples.⁹

It is rather difficult to formally characterize the gap between $E(\varepsilon_t | I_{t-1}^\varepsilon) = 0$ *a.s.* and $E(\varepsilon_t | \varepsilon_{t-j}) = 0$ *a.s.* for all $j > 0$. However, these two hypotheses coincide under some special but important cases. The first case is when $\{\varepsilon_t\}$ is a stationary Gaussian process, which can be a long memory. The second case is when $\{\varepsilon_t\}$ is a Markovian process. This covers both linear and nonlinear Markovian processes. The examples of the latter are a nonlinear autoregressive process $\varepsilon_t = g(\varepsilon_{t-1}) + z_t$ and a bilinear autoregressive process $\varepsilon_t = \alpha \varepsilon_{t-1} + \beta z_t \varepsilon_{t-1} + z_t$, where $\{z_t\} \sim \text{i.i.d. } (0, \sigma^2)$. The third case is when $\{\varepsilon_t\}$ follows an additive-in-mean process: $\varepsilon_t = \alpha_0 + \sum_{j=1}^\infty g_j(\varepsilon_{t-j}) + h_t^{1/2} z_t$, where $g_j(\cdot)$ is not a zero function at least for some lag $j > 0$, and $h_t \equiv h(I_{t-1}^\varepsilon)$ may not be additive. Additive time series processes have attracted considerable interest in the nonparametric literature (*e.g.* Marsry and Tjøstheim (1997), Gao *et al.* (2002), Kim and Linton (2003)).

To reduce the gap between $E(\varepsilon_t | I_{t-1}^\varepsilon) = 0$ *a.s.* and $E(\varepsilon_t | \varepsilon_{t-j}) = 0$ *a.s.* for all $j > 0$, we can extend $f^{(0,1,0)}(\omega, 0, v)$ to a generalized bispectral derivative

$$\left. \frac{\partial}{\partial u} B(\omega_1, \omega_2, u, v_1, v_2) \right|_{u=0} = \frac{1}{(2\pi)^2} \sum_{j=-\infty}^\infty \sum_{l=-\infty}^\infty \sigma_{j,l}^{(1,0,0)}(0, v_1, v_2) e^{-ij\omega_1 - il\omega_2},$$

$$\omega_1, \omega_2 \in [-\pi, \pi],$$

where

$$\sigma_{j,l}^{(1,0,0)}(0, v_1, v_2) \equiv \left. \frac{\partial \sigma_{j,l}(0, v_1, v_2)}{\partial u} \right|_{u=0} = E\{i\varepsilon_t [e^{iv_1 \varepsilon_{t-|j|}} - \varphi(v_1)] [e^{iv_2 \varepsilon_{t-|l|}} - \varphi(v_2)]\}$$

and $\sigma_{j,l}(u, v_1, v_2) \equiv E[(e^{iu\varepsilon_t} - \varphi(u))(e^{iv_1 \varepsilon_{t-|j|}} - \varphi(v_1))(e^{iv_2 \varepsilon_{t-|l|}} - \varphi(v_2))]$ is a generalized third order cumulant. This is equivalent to the use of $E(\varepsilon_t | \varepsilon_{t-j}, \varepsilon_{t-l})$, which was suggested

8. We thank Hidehiko Ichimura for providing this nice example which the generalized spectral derivative $f^{(0,1,0)}(\omega, 0, v)$ will miss.

9. It is well known (*e.g.* Granger and Teräsvirta, 1993, pp. 17, 44) that the class of nonlinear moving averages processes is generally not invertible, and as a consequence, has found little empirical application in practice.

by Tong (1990, p. 219). With $\sigma_{j,l}^{(1,0,0)}(0, v_1, v_2)$, we can detect a larger class of alternatives to $E(\varepsilon_t | I_{t-1}^\varepsilon) = 0$ a.s.¹⁰ For example, it can easily detect the nonlinear moving-average process (3.6). Note that the nonparametric generalized bispectral derivative approach can check many pairs of lags (k, l) , while still avoiding the curse of dimensionality. Nevertheless, in this paper, we focus on $\sigma_j^{(1,0)}(0, v)$ for simplicity.

3.2. Generalized spectral derivative estimation

In the present context, ε_t is not observed. Suppose we have a random sample $\{Y_t\}_{t=1}^T$ which is used to estimate model $g(I_{t-1}, \theta)$. We then obtain the estimated model residual

$$\hat{\varepsilon}_t \equiv Y_t - g(I_{t-1}^\dagger, \hat{\theta}), \quad t = 1, \dots, T, \tag{3.7}$$

where I_{t-1}^\dagger is the information set observed at time $t - 1$, and $\hat{\theta}$ is a \sqrt{T} -consistent estimator for θ_0 . Examples of $\hat{\theta}$ are conditional least squares and quasi-maximum likelihood estimators.

We can estimate $f^{(0,1,0)}(\omega, 0, v)$ by a smoothed kernel estimator

$$\hat{f}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{j=1-T}^{T-1} (1 - |j|/T)^{1/2} k(j/p) \hat{\sigma}_j^{(1,0)}(0, v) e^{-ij\omega}, \quad \omega \in [-\pi, \pi], \tag{3.8}$$

where $\hat{\sigma}_j^{(1,0)}(0, v) = \frac{\partial}{\partial u} \hat{\sigma}_j(u, v)|_{u=0}$, $\hat{\sigma}_j(u, v) = \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v)$, and

$$\hat{\varphi}_j(u, v) = \frac{1}{T - |j|} \sum_{t=|j|+1}^T e^{iu\hat{\varepsilon}_t + iv\hat{\varepsilon}_{t-|j|}}.$$

Here, $p \equiv p(T)$ is a bandwidth, and $k : \mathbb{R} \rightarrow [-1, 1]$ 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. It could be replaced by unity. Under certain conditions, $\hat{f}^{(0,1,0)}(\omega, 0, v)$ is consistent for $f^{(0,1,0)}(\omega, 0, v)$. See Theorem 2 below.

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

$$f_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sigma_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi], \tag{3.9}$$

which can be consistently estimated by

$$\hat{f}_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \hat{\sigma}_0^{(1,0)}(0, v), \quad \omega \in [-\pi, \pi]. \tag{3.10}$$

To test \mathbb{H}_0 , we can compare $\hat{f}^{(0,1,0)}(\omega, 0, v)$ with $\hat{f}_0^{(0,1,0)}(\omega, 0, v)$.

3.3. Tests under conditional heteroscedasticity

There is a growing consensus among economists that the volatilities of most high-frequency economic and financial data are time-varying. It is well known that the asymptotic variances of test statistics for autocorrelations and conditional mean models depend on the type and degree of heteroscedasticity present. Ignoring it will invalidate the limit distributions of test statistics (e.g. Diebold (1986), Lo and MacKinlay (1988), Wooldridge (1990), Whang (1998)). This is also true for our tests. In fact, for our tests, it is also important to take into account other higher order time-varying moments. Some recent studies (e.g. Gallant, Hsieh and Tauchen (1991),

10. Tsay (1986), Hsieh (1989) and Hinich and Patterson (1993) used the third order cumulant $E(\varepsilon_t \varepsilon_{t-j} \varepsilon_{t-l})$ to detect nonlinearity in mean. The measure $\sigma_{j,l}(u, v_1, v_2)$ can detect a wider class of processes than $E(\varepsilon_t \varepsilon_{t-j} \varepsilon_{t-l})$.

Hansen (1994), Harvey and Siddique (1999, 2000), Jondeau and Rockinger (2003) documented time-varying conditional skewness and kurtosis of asset returns. Below, we first propose a test of \mathbb{H}_0 that is robust to conditional heteroscedasticity and other time-varying higher order moments of unknown form. This is one of the most important contributions of our paper in terms of empirical relevance and asymptotic analysis. The asymptotic analysis is nontrivial because of the need to take care of the serial dependence in higher order moments.

Our test statistic that is robust to conditional heteroscedasticity and other time-varying higher order conditional moments of unknown form is given as follows:

$$\hat{M}_1(p) \equiv \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\hat{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) - \hat{C}_1(p) \right] / \sqrt{\hat{D}_1(p)}, \quad (3.11)$$

where $W : \mathbb{R} \rightarrow \mathbb{R}^+$ is a nondecreasing function that weighs sets symmetric about zero equally,

$$\begin{aligned} \hat{C}_1(p) &= \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=j+1}^{T-1} \hat{\varepsilon}_t^2 \int |\hat{\psi}_{t-j}(v)|^2 dW(v), \\ \hat{D}_1(p) &= 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p) k^2(l/p) \int \int \left| \frac{1}{T - \max(j, l)} \right. \\ &\quad \left. \times \sum_{t=\max(j, l)+1}^T \hat{\varepsilon}_t^2 \hat{\psi}_{t-j}(v) \hat{\psi}_{t-l}(v') \right|^2 dW(v) dW(v'), \end{aligned}$$

and $\hat{\psi}_t(v) = e^{iv\hat{\varepsilon}_t} - \hat{\varphi}(v)$, and $\hat{\varphi}(v) = T^{-1} \sum_{t=1}^T e^{iv\hat{\varepsilon}_t}$. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. An example of $W(\cdot)$ is the $N(0, 1)$ CDF, which is commonly used in the characteristic function literature. The factors $\hat{C}_1(p)$ and $\hat{D}_1(p)$ are approximately the mean and the variance of $T \int_{-\pi}^{\pi} |\hat{f}^{(0,1,0)}(\omega, 0, v) - \hat{f}_0^{(0,1,0)}(\omega, 0, v)|^2 d\omega dW(v)$. They have taken into account the impact of conditional heteroscedasticity and other time-varying higher order conditional moments.

3.4. Tests under conditional homoscedasticity

To examine why it is important to take into account the impact of conditional heteroscedasticity and higher order time-varying moments in testing \mathbb{H}_0 , we now derive the generalized spectral tests for \mathbb{H}_0 under conditional homoscedasticity and under i.i.d. for $\{\varepsilon_t\}$, respectively. Suppose $\{\varepsilon_t\}$ is conditionally homoscedastic (*i.e.* $E(\varepsilon_t^2 | I_{t-1}) = \sigma^2$ *a.s.*). Then we can simplify our test statistic as follows:

$$\hat{M}_2(p) \equiv \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\hat{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) - \hat{C}_2(p) \right] / \sqrt{\hat{D}_2(p)}, \quad (3.12)$$

where

$$\begin{aligned} \hat{C}_2(p) &= \hat{s}^2 \int \hat{\sigma}_0(v, -v) dW(v) \sum_{j=1}^{T-1} k^2(j/p), \\ \hat{D}_2(p) &= 2\hat{s}^4 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p) k^2(l/p) \int \int |\hat{\sigma}_{j-l}(v, v')|^2 dW(v) dW(v'), \end{aligned}$$

and $\hat{s}^2 \equiv T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2$ is the sample variance of $\{\varepsilon_t\}_{t=1}^T$. Both the centring and scale factors $\hat{C}_2(p)$ and $\hat{D}_2(p)$ have been simplified, by exploiting the implication of conditional homoscedasticity. The form of $\hat{D}_2(p)$ still takes into account the impact of possibly time-varying third order and higher order moments.

3.5. Tests under the i.i.d. case

Many existing tests for linearity assume i.i.d. for $\{\varepsilon_t\}$ under \mathbb{H}_0 ; some of them further assume i.i.d. $N(0, \sigma^2)$. When $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma^2)$, which implies \mathbb{H}_0 , our test statistic can be further simplified:

$$\hat{M}_3(p) \equiv \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\hat{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) - \hat{C}_3(p) \right] / \sqrt{\hat{D}_3(p)}, \quad (3.13)$$

where $\hat{C}_3(p) = \hat{C}_2(p)$ and $\hat{D}_3(p) = 2\hat{\sigma}^4 \int \int |\hat{\sigma}_0(v, v')|^2 dW(v)dW(v') \sum_{j=1}^{T-2} k^4(j/p)$.

The scale factor $\hat{D}_3(p)$ has been greatly simplified. Interestingly, the $\hat{M}_3(p)$ test derived under conditional homoscedasticity differs from the $\hat{M}_2(p)$ test derived under i.i.d. This is because $\hat{M}_2(p)$ still takes into account possibly time-varying higher order moments (e.g. skewness and kurtosis). We note the fact that the limit distribution of $\hat{M}_3(p)$ test is valid when $\{\varepsilon_t\}$ is i.i.d. does not mean that $\hat{M}_3(p)$ is an omnibus test for i.i.d., because it only checks the serial dependence in mean. To test i.i.d. for $\{\varepsilon_t\}$, the BDS test or Hong and Lee's (2003) test will be suitable. Nevertheless, these i.i.d. tests are not suitable for testing conditional mean specification, because they check the serial dependence in every moment.

4. ASYMPTOTIC DISTRIBUTION

To derive the null limit distribution of our tests, we provide some regularity conditions:

Assumption A1. $\{Y_t\}$ is a strictly stationary time series process such that $\mu_t \equiv E(Y_t | I_{t-1})$ exists a.s., where I_{t-1} is an information set at time $t - 1$ that may contain lagged dependent variables $\{Y_{t-j}, j > 0\}$, lagged shocks $\{\varepsilon_{t-j} \equiv Y_{t-j} - \mu_t, j > 0\}$, as well as current and lagged exogenous variables $\{Z_{t-j}, j \geq 0\}$, with $E(\varepsilon_t^4) \leq C$.

Assumption A2. For each sufficiently large integer q , there exists a strictly stationary process $\{\varepsilon_{q,t}\}$ measurable with respect to the sigma field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}\}$ such that as $q \rightarrow \infty$, $\varepsilon_{q,t}$ is independent of $\{\varepsilon_{t-q-1}, \varepsilon_{t-q-2}, \dots\}$ for each t , $E(\varepsilon_{q,t} | I_{t-1}) = 0$ a.s., $E(\varepsilon_t - \varepsilon_{q,t})^2 \leq Cq^{-\kappa}$ for some constant $\kappa \geq 1$, and $E(\varepsilon_{q,t}^4) \leq C$ for all large q .

Assumption A3. $g(I_{t-1}, \theta)$ is a parametric model for μ_t , where $\theta \in \Theta$ is a finite-dimensional parameter, such that (a) $g(\cdot, \theta)$ is measurable with respect to I_{t-1} for each $\theta \in \Theta$, and (b) with probability one, $g(I_{t-1}, \cdot)$ is continuously twice differentiable with respect to $\theta \in \Theta$, and $E \sup_{\theta \in \Theta} \|\frac{\partial}{\partial \theta} g(I_{t-1}, \theta)\|^4 \leq C$ and $E \sup_{\theta \in \Theta} \|\frac{\partial^2}{\partial \theta \partial \theta'} g(I_{t-1}, \theta)\|^2 \leq C$.

Assumption A4. $\hat{\theta} - \theta_0 = O_P(T^{-1/2})$, where $\theta_0 \equiv p \lim(\hat{\theta}) \in \Theta$.

Assumption A5. Let I_t^\dagger be an observed information set available at time t that may contain some assumed initial values. Then $\lim_{T \rightarrow \infty} \sum_{t=1}^T \{E[\sup_{\theta \in \Theta} |g(I_{t-1}^\dagger, \theta) - g(I_{t-1}, \theta)|]^4\}^{1/4} \leq C$.

Assumption A6. $k : \mathbb{R} \rightarrow [-1, 1]$ is symmetric about 0, and is continuous at 0 and all points except a finite number of points, with $k(0) = 1$ and $|k(z)| \leq C|z|^{-b}$ as $z \rightarrow \infty$ for some $b > 1$.

Assumption A7. $W : \mathbb{R} \rightarrow \mathbb{R}^+$ is nondecreasing and weighs sets symmetric about zero equally, with $\int v^4 dW(v) \leq C$.

Assumption A8. Put $\psi_t(v) \equiv e^{iv\varepsilon_t} - \varphi(v)$ and $\sigma^2 \equiv E(\varepsilon_t^2)$. Then $\{\frac{\partial}{\partial\theta}g(I_t, \theta_0), \varepsilon_t\}$ is a strictly stationary process such that (a) $\sum_{j=1}^{\infty} \|\text{cov}[\frac{\partial}{\partial\theta}g(I_t, \theta_0), \frac{\partial}{\partial\theta}g(I_{t-j}, \theta_0)]\| \leq C$; (b) $\sum_{j=1}^{\infty} \sup_{(u,v) \in \mathbb{R}^2} |\sigma_j(u, v)| \leq C$; (c) $\sum_{j=1}^{\infty} \sup_{v \in \mathbb{R}} \|\text{cov}[\frac{\partial}{\partial\theta}g(I_t, \theta_0), \psi_{t-j}(v)]\| \leq C$; (d) $\sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \sup_{(u,v) \in \mathbb{R}^2} |E[(\varepsilon_t^2 - \sigma^2)\psi_{t-j}(u)\psi_{t-l}(v)]| \leq C$; (e) $\sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} \sup_{v \in \mathbb{R}} \|\kappa_{j,l,\tau}(v)\| \leq C$, where $\kappa_{j,l,\tau}(v)$ is the fourth order cumulant of the joint distribution of the process $\{\frac{\partial}{\partial\theta}g(I_t, \theta_0), \psi_{t-j}(v), \frac{\partial}{\partial\theta}g(I_{t-l}, \theta_0), \psi_{t-\tau}^*(v)\}$.

Assumptions A1 and A2 are regularity conditions on the data generating process (DGP). We impose strict stationarity on $\{Y_t\}$. The existence of the conditional mean μ_t can be ensured by assuming $E(Y_t^2) < \infty$. Assumption A2 is required only under \mathbb{H}_0 . It assumes that the *m.d.s.* $\{\varepsilon_t\}$ can be approximated by a q -dependent *m.d.s.* process $\{\varepsilon_t\}$ arbitrarily well when q is sufficiently large. Horowitz (2003) imposed a similar condition in a different context. Because $\{\varepsilon_t\}$ is a *m.d.s.*, Assumption A2 essentially imposes restrictions on the serial dependence in higher order moments of ε_t . Among other things, it implies ergodicity for $\{\varepsilon_t\}$. It holds trivially when $\{\varepsilon_t\}$ is a q -dependent process with an arbitrarily large but finite order q . It also covers many non-Markovian processes. To appreciate this, we first consider a threshold GARCH(1, 1) error process $\{\varepsilon_t\}$:

$$\begin{cases} \varepsilon_t = h_t^{1/2} z_t, & \{z_t\} \sim \text{i.i.d. } (0, 1), \\ h_t = \gamma + \alpha h_{t-1} + \beta^+ \varepsilon_{t-1}^2 \mathbf{1}(\varepsilon_{t-1} > 0) + \beta^- \varepsilon_{t-1}^2 \mathbf{1}(\varepsilon_{t-1} \leq 0), \end{cases} \quad (4.1)$$

where $\mathbf{1}(\cdot)$ is the indicator function. This was proposed by Glosten, Jagannathan and Runkle (1993). It includes standard GARCH(1, 1) processes if $\beta^+ = \beta^-$. Putting $\xi_t \equiv \gamma + \beta^+ z_t^2 \mathbf{1}(z_t > 0) + \beta^- z_t^2 \mathbf{1}(z_t \leq 0)$, we have $h_t = \gamma + \gamma \sum_{j=1}^{\infty} \prod_{i=1}^j \xi_{t-i}$. Define $\varepsilon_{q,t} \equiv h_{q,t}^{1/2} z_t$, where $h_{q,t} \equiv \gamma + \gamma \sum_{j=1}^q \prod_{i=1}^j \xi_{t-i}$. Then

$$E(\varepsilon_t - \varepsilon_{q,t})^2 = E(h_t^{1/2} - h_{q,t}^{1/2})^2 \leq E(h_t - h_{q,t}) = \gamma \sum_{j=q+1}^{\infty} \prod_{i=1}^j E(\xi_{t-i}) = \frac{\gamma \rho^{q+1}}{1 - \rho},$$

with $\rho \equiv E(\xi_t) = \alpha + \beta^+ + (\beta^- - \beta^+)E[z_t^2 \mathbf{1}(z_t \leq 0)]$. Thus, Assumption A2 holds if $\rho < 1$.

For another example, we consider a general stochastic volatility process for $\{\varepsilon_t\}$:

$$\begin{cases} \varepsilon_t = \exp(\frac{1}{2}h_t) z_t, & \{z_t\} \sim \text{i.i.d. } (0, 1), \\ h_t = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j \eta_{t-j} + \eta_t, & \{\eta_t\} \sim \text{i.i.d. } N(0, \sigma^2), \end{cases} \quad (4.2)$$

where $\sum_{j=1}^{\infty} \alpha_j^2 < \infty$. We do not assume independence between $\{z_t\}$ and $\{\eta_t\}$. Thus, like (4.1), (4.2) can also capture asymmetry in volatility. Now, put $\varepsilon_{q,t} \equiv \exp(\frac{1}{2}h_{q,t}) z_t$, where $h_{q,t} \equiv \alpha_0 + \sum_{j=1}^q \alpha_j z_{t-j} + z_t$. Then we have

$$\begin{aligned} E(\varepsilon_t - \varepsilon_{q,t})^2 &= E\{e^{\frac{1}{2}h_{q,t}} [e^{\frac{1}{2}(h_t - h_{q,t})} - 1] z_t\}^2 \leq \{E[e^{2h_{q,t}} (e^{\frac{1}{2}(h_t - h_{q,t})} - 1)^4]\}^{\frac{1}{2}} [E(z_t^4)]^{\frac{1}{2}} \\ &= [E(z_t^4)]^{\frac{1}{2}} [E(e^{2h_{q,t}})]^{\frac{1}{2}} \{E[(e^{\frac{1}{2}(h_t - h_{q,t})} - 1)^4]\}^{\frac{1}{2}} \leq C \sum_{j=q+1}^{\infty} \alpha_j^2 \end{aligned}$$

provided $E(z_t^4) < \infty$, where we have made use of the fact that the $N(0, \sigma^2)$ moment generating function is $E(e^{\lambda \eta_t}) = \exp(\frac{1}{2}\sigma^2 \lambda^2)$ for $\lambda \in \mathbb{R}$ and the inequality that $|e^x - 1| \leq 2|x|$ for small x . Thus, Assumption A2 holds if $\sum_{j=q+1}^{\infty} \alpha_j^2 \leq Cq^{-\kappa}$. A sufficient condition is $\alpha_j \leq Cj^{-(\kappa+1)}$. This rules out long-memory stochastic volatility processes given $\kappa \geq 1$, because $\sum_{j=-\infty}^{\infty} \text{cov}(h_t, h_{t-j}) = (\sum_{j=1}^{\infty} \alpha_j)^2 < \infty$.

Assumption A3 is a standard condition on the conditional mean model $g(I_{t-1}, \theta)$. For a linear model $g(I_{t-1}, \theta) = X_t'\theta$, where $X_t \in I_{t-1}$ is finite-dimensional, it suffices if $E\|X_t\|^4 \leq C$. Assumption A3 covers many stationary nonlinear time series conditional mean models, such as bilinear autoregressive, exponential autoregressive, nonlinear moving average, Markov regime-switching, smooth transition and Poisson jump models. It also covers threshold autoregression models with known thresholds. An example is the class of self-exciting autoregressive threshold models for the U.S. economy, where recession and expansion are defined as the GDP growth rate being larger or smaller than zero (Tiao and Tsay (1994), Potter (1995)). However, Assumption A3 rules out the autoregressive threshold models with unknown thresholds considered in Hansen (2000), where $g(I_{t-1}, \theta)$ is not continuous in threshold parameters. We conjecture that our tests are applicable to these models under additional conditions, but we do not attempt to justify this here, as it is beyond the scope of this paper.

Assumption A4 requires a \sqrt{T} -consistent estimator $\hat{\theta}$, which may not be asymptotically most efficient. It can be a conditional least squares estimator or a conditional quasi-maximum likelihood estimator. Also, we do not need to know the asymptotic expansion of $\hat{\theta}$, because the sampling variation in $\hat{\theta}$ does not affect the limit distributions of $\hat{M}_a(p)$. These features are similar in spirit to Wooldridge's (1990) heteroscedasticity-robust modified moment-based tests. Assumption A5 is a condition on the truncation of information set I_{t-1} , which usually contains information dating back to the very remote past and so may not be observable. Because of the truncation, one may have to assume some initial values in estimating the model $g(I_{t-1}, \theta)$; Assumption A5 ensures that the use of initial values, if any, has no impact on the limit distribution of $\hat{M}_a(p)$. For instance, consider an ARMA(1, 1) model:

$$g(I_{t-1}, \theta) = \alpha Y_{t-1} + \beta \varepsilon_{t-1},$$

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

$$\begin{aligned} & \sum_{t=1}^T \{E[\sup_{\theta \in \Theta} |g(I_{t-1}^\dagger, \theta) - g(I_{t-1}, \theta)|^4]^{\frac{1}{4}}\} \\ &= \sum_{t=1}^T \left\{ E \left[\sup_{\theta \in \Theta} \left| \beta \sum_{j=t-1}^\infty \alpha^j \varepsilon_{t-j-1} - \beta \alpha^{t-1} \bar{\varepsilon}_0 \right|^4 \right]^{\frac{1}{4}} \right\} \\ &\leq \bar{\beta} \sum_{t=1}^T \left\{ E \left[\sup_{\alpha} \left| \alpha^{t-1} \left(\sum_{l=0}^\infty \alpha^l \varepsilon_{t-l} - \bar{\varepsilon}_0 \right) \right|^4 \right]^{\frac{1}{4}} \right\} \\ &\leq \bar{\beta} \sum_{t=1}^T |\bar{\alpha}|^{t-1} \left[[E(\varepsilon_0^4)]^{\frac{1}{4}} \sum_{l=0}^\infty \bar{\alpha}^l + [E|\bar{\varepsilon}_0^4|]^{\frac{1}{4}} \right] \leq C. \end{aligned}$$

Assumption A6 is a regularity condition on the kernel $k(\cdot)$. It includes all commonly used kernels in practice. The condition of $k(0) = 1$ ensures that the asymptotic bias of the smoothed kernel estimator $\hat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.8) vanishes as $T \rightarrow \infty$. The tail condition on $k(\cdot)$ requires that $k(z)$ decays to zero sufficiently fast as $|z| \rightarrow \infty$. It implies $\int_0^\infty (1+z)k^2(z)dz < \infty$. For kernels with bounded support, such as the Bartlett and Parzen kernels, $b = \infty$. For the Daniell and quadratic-spectral kernels, $b = 1$ and 2 , respectively. These two kernels have unbounded support, and thus all $T - 1$ lags contained in the sample are used in constructing our test statistics. Assumption A7 is a condition on the weighting function $W(\cdot)$ for the transform parameter v . It is satisfied by the CDF of any symmetric continuous distribution with a finite fourth moment. Finally, Assumption A8 provides some covariance and fourth order cumulant conditions on $\{\frac{\partial}{\partial \theta} g(I_{t-1}, \theta_0), \varepsilon_t\}$, which restrict the degree of the serial dependence in $\{\frac{\partial}{\partial \theta} g(I_{t-1}, \theta_0), \varepsilon_t\}$. These conditions can be ensured by imposing more restrictive mixing and moment conditions

on the process $\{\frac{\partial}{\partial \theta} g(I_{t-1}, \theta_0), \varepsilon_t\}$. However, we do not do so to cover a sufficiently general class of DGPs.

We now state the asymptotic distribution of the $\hat{M}_a(p)$ tests under \mathbb{H}_0 .

Theorem 1. *Suppose Assumptions A1–A8 hold, and $p = cT^\lambda$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ and $0 < c < \infty$. (i) $\hat{M}_1(p) \xrightarrow{d} N(0, 1)$ under \mathbb{H}_0 . (ii) If in addition $E(\varepsilon_t^2 | I_{t-1}) = \sigma^2$ a.s., then $\hat{M}_2(p) \xrightarrow{d} N(0, 1)$ under \mathbb{H}_0 . (iii) If $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma^2)$, then $\hat{M}_3(p) \xrightarrow{d} N(0, 1)$.*

As an important feature of $\hat{M}_a(p)$, the use of the *estimated* model residuals $\{\hat{\varepsilon}_t\}$ in place of the true unobservable errors $\{\varepsilon_t\}$ has no impact on the limit distribution of $\hat{M}_a(p)$. One can proceed as if the true parameter value θ_0 were known and equal to $\hat{\theta}$. The reason is that the convergence rate of the parametric parameter estimator $\hat{\theta}$ to θ_0 is faster than that of the nonparametric kernel estimator $\hat{f}^{(0,1,0)}(\omega, 0, v)$ to $f^{(0,1,0)}(\omega, 0, v)$. Consequently, the limit distribution of $\hat{M}_a(p)$ is solely determined by $\hat{f}^{(0,1,0)}(\omega, 0, v)$, and replacing θ_0 by $\hat{\theta}$ has no impact asymptotically. This delivers a convenient procedure, because no specific estimation method for θ_0 is required. Of course, parameter estimation uncertainty in $\hat{\theta}$ may have impact on the small sample distribution of $\hat{M}_a(p)$. In small samples, one can use a bootstrap procedure similar to Hansen (1996) to obtain more accurate levels of the tests.

Because parameter estimation uncertainty in $\hat{\theta}$ has no impact on the limit distribution of $\hat{M}_1(p)$, $\hat{M}_1(p)$ can be used to test the *m.d.s.* hypothesis for observed raw data with conditional heteroscedasticity of unknown form. No modification to the test statistic $\hat{M}_1(p)$ or its limit distribution is needed. Lobato (2002) and Park and Whang (2003) proposed some nonparametric tests of the *m.d.s.* for observed raw data using the conditioning indicator function. They also allowed for conditional heteroscedasticity, and Park and Whang (2003) allowed for nonstationary conditioning variables. However, these tests only check a fixed lag order. Moreover, their limit distributions depend on the DGP and cannot be tabulated; resampling methods have to be used to obtain critical values on a case-by-case basis.

5. ASYMPTOTIC POWER

Our tests are derived without assuming an alternative model. To gain insight into the nature of the alternatives that our tests are able to detect, we now examine the asymptotic behaviour of $\hat{M}_a(p)$ under \mathbb{H}_A in (2.3). For this purpose, we impose a condition on the serial dependence in $\{\varepsilon_t\}$.

Assumption A9. $\sum_{j=1}^{\infty} \sup_{v \in \mathbb{R}} |\sigma_j^{(1,0)}(0, v)| \leq C$.

Theorem 2. *Suppose Assumptions A1 and A3–A9 hold, and $p = cT^\lambda$ for $0 < \lambda < \frac{1}{2}$ and $0 < c < \infty$. Then as for $a = 1, 2, 3$,*

$$\begin{aligned} (p^{1/2}/T)\hat{M}_a(p) &\xrightarrow{P} \left[2D \int_0^\infty k^4(z) dz \right]^{-1/2} \pi \int \int_{-\pi}^\pi |f^{(0,1,0)}(\omega, 0, v) \\ &\quad - f_0^{(0,1,0)}(\omega, 0, v)|^2 d\omega dW(v) \\ &= \left[2D \int_0^\infty k^4(z) dz \right]^{-1/2} \sum_{j=1}^{\infty} \int |\sigma_j^{(1,0)}(0, v)|^2 dW(v) \end{aligned}$$

where $D \equiv \sigma^4 \sum_{j=-\infty}^{\infty} \int \int |\sigma_j(u, v)|^2 dW(u) dW(v) = 2\pi\sigma^4 \int \int_{-\pi}^\pi |f(\omega, u, v)|^2 d\omega dW(u) dW(v)$.

The constant D takes into account the impact of the serial dependence in conditioning variables $\{e^{iv\varepsilon_{t-j}}, j > 0\}$, which generally exists even under \mathbb{H}_0 , due to the presence of the serial dependence in the conditional variance and higher order moments of $\{\varepsilon_t\}$. This differs from the i.i.d. case, where $D = \sigma^4 \int \int |\sigma_0(v, v')|^2 dW(v) dW(v')$ depends only on the marginal distribution of ε_t .

Suppose the autoregression function $E(\varepsilon_t | \varepsilon_{t-j}) \neq 0$ at some lag $j > 0$. Then we have $\int |\sigma_j^{(1,0)}(0, v)|^2 dW(v) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded support on \mathbb{R} . As a consequence, $\lim_{T \rightarrow \infty} P[M_a(p) > C(T)] = 1$ for any constant $C(T) = o(T/p^{\frac{1}{2}})$. Therefore, $\hat{M}_a(p)$ has asymptotic unit power at any given significance level, whenever $E(\varepsilon_t | \varepsilon_{t-j})$ is nonzero at some lag $j > 0$.¹¹ We thus expect that $\hat{M}_a(p)$ has relatively omnibus power against a wide variety of linear and nonlinear alternatives with unknown lag structure, as is confirmed in our simulation below. It should be emphasized that the omnibus power property does not mean that $\hat{M}_a(p)$ is more powerful than *any* other existing tests against *every* alternative. In fact, just because $\hat{M}_a(p)$ has to take care of a wide range of possible alternatives, it may be less powerful against certain specific alternatives than the parametric test in finite samples. Nevertheless, the main advantage of $\hat{M}_a(p)$ is that it can eventually detect all possible model misspecifications that render $E(\varepsilon_t | \varepsilon_{t-j})$ nonzero at some lag $j > 0$. This avoids the blindness of searching for different alternatives when one has no prior information.

Because the existing tests for linearity in mean only consider a fixed order lag, they can easily miss misspecifications at higher lag orders. Of course, these tests could be used to check a large number of lags when a large sample is available. However, they are not expected to be powerful against many alternatives of practical importance, due to the loss of a large number of degrees of freedom. This power loss is greatly alleviated for our tests due to the role played by $k^2(\cdot)$. Most nonuniform kernels discount higher order lags. This enhances good power against the alternatives whose serial dependence decays to zero as lag order j increases. Thus, our tests 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.

Once the model $g(I_{t-1}, \theta)$ is rejected by our omnibus test $\hat{M}_a(p)$, one may like to go further to explore possible sources of model misspecification in mean. For this purpose, we can further differentiate the generalized spectral derivative $f^{(0,1,0)}(\omega, 0, v)$ with respect to v and construct a sequence of tests similar in spirit to our $\hat{M}_a(p)$ tests. In particular, the derivatives $\sigma_j^{(1,l)}(0, 0)$ with $l = 1, 2, 3, 4$ yield $\text{cov}(\varepsilon_t, \varepsilon_{t-j})$, $\text{cov}(\varepsilon_t, \varepsilon_{t-j}^2)$, $\text{cov}(\varepsilon_t, \varepsilon_{t-j}^3)$ and $\text{cov}(\varepsilon_t, \varepsilon_{t-j}^4)$, respectively. Tests based on these derivatives can thus tell us whether there exists linear correlation, ARCH-in-mean, skewness-in-mean or kurtosis-in-mean effects, respectively. ARCH-in-mean effects are important in finance (Engle *et al.*, 1987), and the recent literature also has documented time-varying skewness and kurtosis and their economic relevance in asset pricing (Harvey and Siddique, 1999, 2000).

The $\hat{M}_a(p)$ tests may not be very powerful in finite samples under some scenarios, because no direct use of lagged dependent variables and exogenous variables (if any) has been made. This is the case particularly when μ_t can be more effectively characterized by some vector $Z_t \in I_{t-1}$ than by the ε_{t-j} , where Z_t may contain lagged dependent variables and/or exogenous variables. To improve power, we may extend our approach to include Z_t as the conditioning variables.

11. Since $\int \int_{-\pi}^{\pi} |f^{(0,1,0)}(\omega, 0, v) - f_0^{(0,1,0)}(\omega, 0, v)|^2 d\omega dW(v)$ is strictly positive whenever $E(\varepsilon_t | \varepsilon_{t-j}) \neq 0$ for some lag $j > 0$, upper-tailed asymptotic $N(0, 1)$ critical values (e.g. 1.645 at the 5% level) should be used.

Put $X_t \equiv (Z_t', \varepsilon_t)'$, and define

$$\sigma_{\varepsilon X, j}^{(1,0)}(0, v) \equiv \text{cov}(i\varepsilon_t, e^{iv'X_{t-|j|}}), \quad v \in \mathbb{R}^{d+1},$$

where d is the dimension of Z_t .¹² The associated test statistic of \mathbb{H}_0 that is robust to conditional heteroscedasticity and higher order moments of unknown form is given as follows:

$$\hat{M}_{\varepsilon X}(p) \equiv \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int_{\mathbb{R}^{d+1}} |\hat{\sigma}_{\varepsilon X}^{(1,0)}(0, v)|^2 d\mathcal{W}(v) - \hat{C}_{\varepsilon X, 1}(p) \right] / \sqrt{\hat{D}_{\varepsilon X, 1}(p)}, \quad (5.1)$$

where $\hat{\sigma}_{\varepsilon X}^{(1,0)}(0, v) = \frac{\partial}{\partial u} \hat{\sigma}_{\varepsilon X}(u, v)|_{u=0}$, $\hat{\sigma}_{\varepsilon X}(u, v) = \hat{\varphi}_{\varepsilon X, j}(u, v) - \hat{\varphi}_{\varepsilon X, j}(u, 0)\hat{\varphi}_{\varepsilon X, j}(0, v)$, $\hat{\varphi}_{\varepsilon X, j}(u, v) \equiv (T-j)^{-1} \sum_{t=|j|+1}^T e^{iu\hat{\varepsilon}_t + iv'\hat{X}_{t-|j|}}$, and $\hat{X}_t \equiv (Z_t', \hat{\varepsilon}_t)'$. In addition, $\hat{C}_{\varepsilon X, 1}(p)$ and $\hat{D}_{\varepsilon X, 1}(p)$ are defined in the same way as $\hat{C}_1(p)$ and $\hat{D}_1(p)$ in (3.11) with $\hat{\psi}_t(v)$ replaced by $\hat{\psi}_{X, t-j}(v) \equiv e^{iv'\hat{X}_{t-|j|}} - \hat{\varphi}_X(v)$, where $\hat{\varphi}_X(v) = T^{-1} \sum_{t=1}^T e^{iv'\hat{X}_t}$. For simplicity, we can use a product weighting function $\mathcal{W}(v) = \prod_{j=1}^{d+1} W(v_j)$, where $W(\cdot)$ satisfies Assumption A7. The test statistic $\hat{M}_{\varepsilon X}(p)$ involves $(d+1)$ - and $2(d+1)$ -dimensional numerical integrations, which can be evaluated by simulation when d is large.

Under suitable conditions, we can show that $\hat{M}_{\varepsilon X}(p) \xrightarrow{p} N(0, 1)$ under \mathbb{H}_0 , and

$$(p^{1/2}/T)\hat{M}_{\varepsilon X}(p) \xrightarrow{p} \left[2D_X \int_0^\infty k^4(z)dz \right]^{-1/2} \sum_{j=1}^\infty \int |\sigma_{\varepsilon X, j}^{(1,0)}(0, v)|^2 d\mathcal{W}(v)$$

under \mathbb{H}_A , where $D_X \equiv \sigma^4 \sum_{j=1}^\infty \int \int |\sigma_{X X, j}(u, v)|^2 d\mathcal{W}(u)d\mathcal{W}(v)$ and $\sigma_{X X, j}(u, v) = \text{cov}(e^{iu'X_t}, e^{iv'X_{t-|j|}})$. Thus, this test can detect all model misspecifications in mean that render $E(\varepsilon_t | X_{t-j}) \equiv E(\varepsilon_t | Z_{t-j}, \varepsilon_{t-j}) \neq 0$ for some $j > 0$.

A plausible alternative approach to testing \mathbb{H}_0 in (2.2) is to consider a test based on the statistic

$$\sum_{j=1}^{T-1} k^2(j/p)(T-j) \left[\frac{1}{T-j} \sum_{t=j+1}^T \hat{m}_j^2(\hat{\varepsilon}_{t-j}) \right], \quad (5.2)$$

where $\hat{m}_j(\cdot)$ is a smoothed nonparametric estimator for the autoregression function $E(\varepsilon_t | \varepsilon_{t-j})$ at lag j . This test also avoids the curse of dimensionality and can detect the same class of global alternatives to \mathbb{H}_0 as the $\hat{M}_a(p)$ tests. Moreover, the estimator $\hat{m}_j(\cdot)$ is often of more direct interest in practice. Nevertheless, unlike $\hat{\sigma}_j(0, v)$, $\hat{m}_j(\cdot)$ involves a smoothing parameter, say a bandwidth h_j (for simplicity, one can set $h_j = h \equiv h(T)$ for all $j > 0$). Closely associated with the choice of the bandwidth h , the test based on (5.2) is expected to have poor levels even for rather large sample sizes, as was documented in Hjellvik and Tjøstheim (1996), who used a statistic similar to (5.2) to test linearity for observed raw data, with a fixed p and the truncated (uniform) kernel $k(z) = \mathbf{1}(|z| \leq 1)$. Hjellvik and Tjøstheim (1996) pointed out that in a Taylor series expansion of (5.2), asymptotically negligible higher order terms depend on h and are nearly of the same order of magnitude as the dominating term that determines the asymptotic distribution. As a consequence, to retain the dominating term and to neglect the others will give poor approximations in finite samples.¹³

12. The Fourier transform, $f_{\varepsilon X}^{(0,1,0)}(\omega, 0, v) = (2\pi)^{-1} \sum_{j=-\infty}^\infty \sigma_{\varepsilon X, j}^{(1,0)}(0, v) e^{-ij\omega}$, $\omega \in [-\pi, \pi]$, may be viewed as a generalized cross-spectrum between $\{\varepsilon_t\}$ and $\{X_t\}$.

13. Of course, one can use bootstrap to obtain correct levels for this test. But this will complicate the implementation of the test. In particular, the naive bootstrap is not suitable, because there exists serial dependence in the variance and higher order moments.

6. DATA-DRIVEN LAG ORDER

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

Assumption A10. For any $x, y \in \mathbb{R}$, $|k(x) - k(y)| \leq C|x - y|$ for some constant $C < \infty$.

Theorem 3. Suppose Assumptions A1–A8 and A10 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 A6, and p is a nonstochastic bandwidth with $p = cT^\lambda$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ and $0 < c < \infty$. Then (i) $\hat{M}_1(\hat{p}) - \hat{M}_1(p) \xrightarrow{p} 0$ and $\hat{M}_1(\hat{p}) \xrightarrow{d} N(0, 1)$ under \mathbb{H}_0 . (ii) If in addition $E(\varepsilon_t^2 | I_{t-1}) = \sigma^2$ a.s., then $\hat{M}_2(\hat{p}) - \hat{M}_2(p) \xrightarrow{p} 0$ and $\hat{M}_2(\hat{p}) \xrightarrow{d} N(0, 1)$ under \mathbb{H}_0 . (iii) If $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma^2)$, then $\hat{M}_3(\hat{p}) - \hat{M}_3(p) \xrightarrow{p} 0$ and $\hat{M}_3(\hat{p}) \xrightarrow{d} N(0, 1)$.

Thus, as long as \hat{p} converges to p sufficiently fast, the use of \hat{p} instead of p has no impact on the limit distribution of $\hat{M}_a(\hat{p})$. 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 similar to Hong (1999, Theorem 2.2). It minimizes an asymptotic integrated mean squared error (IMSE) criterion for the estimator $\hat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.8). Consider some “pilot” generalized spectral derivative estimators based on a preliminary bandwidth \bar{p} :

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

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

where the kernel $\bar{k}(\cdot)$ need not be the same as the kernel $k(\cdot)$ used in (3.8). For example, $\bar{k}(\cdot)$ can be the Bartlett kernel while $k(\cdot)$ is the Parzen kernel. Note that $\bar{f}^{(0,1,0)}(\omega, 0, v)$ is an estimator for $f^{(0,1,0)}(\omega, 0, v)$ and $\bar{f}^{(q,1,0)}(\omega, 0, v)$ is an estimator for the generalized spectral derivative

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

For the kernel $k(\cdot)$, suppose there exists some $q \in (0, \infty)$ such that $0 < k^{(q)} \equiv \lim_{z \rightarrow 0} \frac{1 - k(z)}{|z|^q} < \infty$. Then we define the plug-in bandwidth

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

where the tuning parameter estimator

$$\begin{aligned} \hat{c}_0 &\equiv \left[\frac{2q(k^{(q)})^2 \int_{-\pi}^{\pi} |\bar{f}^{(q,1,0)}(\omega, 0, v)|^2 d\omega dW(v)}{\int_{-\infty}^{\infty} k^2(z) dz \int_{-\pi}^{\pi} [\int \bar{f}^{(0,1,0)}(\omega, v, -v) dW(v)]^2 d\omega} \right]^{\frac{1}{2q+1}} \\ &= \left[\frac{2q(k^{(q)})^2 \sum_{j=1-T}^{T-1} (T - |j|) \bar{k}^2(j/\bar{p}) |j|^{2q} \int |\hat{\sigma}_j^{(1,0)}(0, v)|^2 dW(v)}{\int_{-\infty}^{\infty} k^2(z) dz \sum_{j=1-T}^{T-1} (T - |j|) \bar{k}^2(j/\bar{p}) \hat{R}(j) \int \hat{\sigma}_j(v, -v) dW(v)} \right]^{\frac{1}{2q+1}}, \end{aligned}$$

and $\hat{R}(j) \equiv (T - |j|)^{-1} \sum_{t=|j|+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-|j|}$.¹⁴ Note that $\hat{\rho}_0$ is real-valued. One can take its integer part, and the impact of integer-clipping is expected to be negligible.

The data-driven $\hat{\rho}_0$ in (6.4) involves the choice of a preliminary bandwidth \bar{p} , which can be fixed or grow with the sample size T . If \bar{p} is fixed, $\hat{\rho}_0$ still generally grows at rate $T^{\frac{1}{2q+1}}$ under \mathbb{H}_A , but \hat{c}_0 does not converge to the optimal tuning constant c_0 (say) that minimizes the IMSE of $\hat{f}^{(0,1,0)}(\omega, 0, v)$. This is a parametric plug-in method. Alternatively, following Hong (1999), we can show that when \bar{p} grows with T properly, the data-driven bandwidth $\hat{\rho}_0$ in (6.4) will minimize an asymptotic IMSE of $\hat{f}^{(0,1,0)}(\omega, 0, v)$. The choice of \bar{p} is somewhat arbitrary, but we expect that it is of secondary importance. This is confirmed in our simulation and empirical application below.¹⁵

We emphasize that the data-driven \hat{p} based on the IMSE criterion generally will not maximize the power of $\hat{M}_a(p)$. A more sensible alternative would be to develop a data-driven \hat{p} using a power criterion, or a criterion that trades off level distortion and power loss. This will necessitate higher order asymptotic analysis and is beyond the scope of this paper. We are content with the IMSE criterion here. Our simulation experience suggests that the power of our tests seems to be relatively flat in the neighbourhood of the optimal lag order that maximizes the power, and $\hat{\rho}_0$ in (6.4) performs reasonably well in finite samples. Nevertheless, the issue of the optimal data-driven \hat{p} for our tests is far from being resolved from a theoretical perspective.

7. MONTE CARLO EVIDENCE

We now investigate the finite sample performance of our $\hat{M}_a(p)$ tests. While our tests can be used to test nonlinear conditional mean models, we focus on testing linearity in mean, which has attracted a lot of attention in the literature. Because our tests are derived without specifying an alternative, we will compare them with a number of popular linearity tests of similar spirit, namely those of Tsay (1986), White (1989) and Hamilton (2001). The limit distributions of these tests are derived under conditional homoscedasticity or i.i.d. $N(0, \sigma^2)$. Their robust versions under conditional heteroscedasticity of unknown form have not been available.

7.1. Simulation design

7.1.1. Level. To examine the levels of the tests under \mathbb{H}_0 , we consider the following DGPs:

$$\text{DGP S.1 [AR(1)-i.i.d. (0, 1)]: } Y_t = 0.5Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N(0, 1),$$

$$\text{DGP S.2 [AR(1)-ARCH(1)]: } \begin{cases} Y_t = 0.5Y_{t-1} + \varepsilon_t, \\ \varepsilon_t = h_t^{1/2} z_t, \quad h_t = 0.43 + 0.57\varepsilon_{t-1}^2, \\ z_t \sim \text{i.i.d. } N(0, 1). \end{cases}$$

14. For the $\hat{M}_{\varepsilon X}(p)$ test in (5.1), we need to modify \hat{c}_0 as follows: $\hat{\sigma}_j^{(1,0)}(0, v)$ in the numerator of \hat{c}_0 should be replaced by $\hat{\sigma}_{\varepsilon X, j}^{(1,0)}(0, v)$ as used in (5.1), and $\hat{\sigma}_j(v, -v)$ in the denominator should be replaced by $\hat{\sigma}_{XX, j}(v, -v) = \hat{\varphi}_{XX, j}(v, -v) - \hat{\varphi}_{XX, j}(v, 0)\hat{\varphi}_{XX, j}(0, -v)$, where $\hat{\varphi}_{XX, j}(u, v) = (T - |j|)^{-1} \sum_{t=|j|+1}^T e^{iu' \hat{X}_t + iv' \hat{X}_{t-|j|}}$, and $\hat{X}_t \equiv (Z_t', \hat{\varepsilon}_t)'$. Moreover, the univariate weighting function $W(v)$ should be replaced by the product weighting function $\mathcal{W}(v) = \prod_{j=1}^{d+1} W(v_j)$, where $v = (v_1, \dots, v_{d+1})'$.

15. The tuning parameter estimator \hat{c}_0 will converge to zero under \mathbb{H}_0 . To ensure $\hat{\rho}_0 \rightarrow \infty$, we can use the formula (6.4) supplemented with a slow-growing lower bound (say $\ln T$) such that $\tilde{\rho}_0 = \max(\ln T, \hat{\rho}_0)$. The choice of the slow-growing lower bound $\ln T$ is arbitrary, but it will not affect the IMSE-optimal rate $\hat{\rho}_0$ under \mathbb{H}_A , when T is sufficiently large.

Under these DGPs, the linear AR(1) model

$$g(I_{t-1}, \theta) = \theta_1 + \theta_2 Y_{t-1} \tag{7.1}$$

is correctly specified for $E(Y_t | I_{t-1}) = 0.5Y_{t-1}$. The parameter $\theta_0 \equiv (0, 0.5)'$ can be estimated consistently by the OLS estimator $\hat{\theta}$. The model error $\{\varepsilon_t(\theta_0)\}$ is conditionally homoscedastic under DGP S.1; all tests considered are asymptotically valid under this DGP. Under DGP S.2, $\{\varepsilon_t(\theta_0)\}$ is conditionally heteroscedastic; only $\hat{M}_1(\hat{\rho}_0)$ has a valid limit distribution. This allows us to examine the importance of taking into account conditional heteroscedasticity. We have chosen parameter values in DGP S.2 such that $E[\varepsilon_t^4(\theta_0)] < \infty$.¹⁶ To examine the level, we consider three sample sizes: $T = 100, 250$ and 500 .

7.1.2. Power. Next, we examine the power of the tests for neglected nonlinearity or dynamic misspecification in mean (*i.e.* lag order misspecification). Because the other tests for linearity in mean are not valid under conditional heteroscedasticity, we will focus on homoscedastic errors in power comparison. We consider the following DGPs:

DGP P.1 [Bilinear(1)]: $Y_t = 0.5Y_{t-1} + 0.6Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$,

DGP P.2 [NMA(1)]: $Y_t = 0.5Y_{t-1} - 0.6\varepsilon_{t-1}^2 + \varepsilon_t$,

DGP P.3 [EXP-AR(1)]: $Y_t = 0.5Y_{t-1} + 10Y_{t-1} \exp(-Y_{t-1}^2) + \varepsilon_t$,

DGP P.4 [SETAR(1)]: $Y_t = \begin{cases} 0.5Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \leq 0, \\ -0.5Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} > 0, \end{cases}$

DGP P.5 [STAR(1)]: $Y_t = 1 - 0.5Y_{t-1} - (4 + 0.4Y_{t-1})G(\alpha Y_{t-1}) + \varepsilon_t$,
 where $G(z) = [1 + \exp(-z)]^{-1}$,

DGP P.6 [ARMA(1, 1)]: $Y_t = 0.5Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$,

DGP P.7 [NMA(5)]: $Y_t = 0.5Y_{t-1} + \sum_{j=1}^5 0.5^j \varepsilon_{t-j}^2 + \varepsilon_t$,

DGP P.8 [SIGN AR(6)]: $Y_t = \mathbf{1}(Y_{t-6} > 0) - \mathbf{1}(Y_{t-6} < 0) + \varepsilon_t$,

where $\{\varepsilon_t\}$ is i.i.d. $N(0, 1)$.

DGPs P.1–P.5 are taken from Tsay (2001). They cover bilinear autoregressive, nonlinear MA, exponential autoregressive, self-exciting threshold, and smooth transition threshold processes. These DGPs allow us to examine the power of the tests against various neglected nonlinearities in mean. DGP P.6 is a linear ARMA(1, 1) process, under which the AR(1) model in (7.1) suffers from dynamic misspecification rather than neglected nonlinearity in mean. Note that DGPs P.1–P.6 are all first order dynamic processes. Our tests, which employ several lags, may not be expected to be the most powerful against them. To highlight our approach, we include DGPs P.7 and P.8. DGP P.7 is an extended nonlinear MA process up to lag 5, with declining coefficients. Declining weights reflect the stylized fact that economic behaviours are usually more influenced by the recent past events than the remote past events. DGP P.8 is a sign autoregressive process with a time delay of six periods. Time delay is an important feature in nonlinear time series (Tong (1990), Granger and Teräsvirta (1993, p. 8)). After removing the best least square approximation to each of DGPs P.1–P.8, we find that for all these DGPs, the generalized spectral derivative $m(\omega)$ in (3.5) of the model error $\varepsilon_t \equiv \varepsilon_t(\theta_0)$, where $\theta_0 = p \lim \hat{\theta}$, is not flat over the frequency ω .¹⁷

16. We also consider a GARCH process with an infinite fourth order moment. The level performance of the generalized spectral derivative tests is similar.

17. In checking whether $m(\omega)$ is flat, we use a numerical method. We first generate a data of 10,000 observations and obtain the OLS residuals of the AR(1) model in (7.1) fitted to the simulated data. Then we obtain the kernel estimator

Thus, our tests are expected to be able to detect model misspecifications under all these DGPs for the sample size T sufficiently large, in light of Theorem 2. We will examine how powerful our tests are relative to the other tests for two sample sizes: $T = 100$ and 250 , respectively.

7.2. Computation of test statistics

To compute the test statistics of Tsay (1986), White (1989) and Hamilton (2001), one has to determine how many lags (p) should be used in the test vector $Z_t \equiv (Y_{t-1}, \dots, Y_{t-p})'$. We choose $p = 1, 2$ and 5 for these tests.¹⁸ Tsay's (1986) test statistic is computed as follows: (i) regress Y_t on 1 and Y_{t-1} and save the estimated residuals $\{\hat{\varepsilon}_t\}$; (ii) let H_t be the vector containing $\frac{1}{2}p(p+1)$ cross-product terms of the components of Z_t , in the form $Y_{t-j}Y_{t-k}$, where $j \geq k$ and $j, k = 1, \dots, p$. Then regress H_t on 1 and Y_{t-1} and save the vector-valued residuals $\{\hat{\xi}_t\}$; (iii) regress $\hat{\varepsilon}_t$ on $\hat{\xi}_t$ and save the estimated residuals $\{\hat{v}_t\}$; (iv) compute Tsay's statistic

$$\text{TSAY}(p) = \frac{\left(\sum_{t=1}^T \hat{\xi}_t \hat{\varepsilon}_t\right)' \left(\sum_{t=1}^T \hat{\xi}_t \hat{\xi}_t'\right)^{-1} \left(\sum_{t=1}^T \hat{\xi}_t \hat{\varepsilon}_t\right) / \frac{1}{2}p(p+1)}{\left(\sum_{t=1}^T \hat{v}_t^2\right) / \left(T - p - \frac{1}{2}p(p+1) - 1\right)},$$

and compare it to the $F[\frac{1}{2}p(p+1), T - \frac{1}{2}p(p+3) - 1]$ distribution.¹⁹

White's (1989) neural network test statistic is computed as follows: (i) regress Y_t on 1 and Y_{t-1} and save the estimated residuals $\{\hat{\varepsilon}_t\}$; (ii) construct the test function $\psi_{jt} \equiv \psi(\gamma_{j0} + \gamma_j' Z_t)$, $j = 1, \dots, q$, where $\psi(\cdot)$ is the activation function, $\Gamma_j \equiv (\gamma_{j0}, \gamma_j)'$ is randomly drawn from a uniform distribution on $[-2, 2]^{p+1}$, and q is a pre-specified number of hidden units. Following Lee *et al.* (1993), we scale $\{Y_t\}$ to $[0, 1]$, using the logistic function $\psi(u) = [1 + \exp(-u)]^{-1}$, and set $q = 10$; (iii) compute the principal components of $\psi_t \equiv (\psi_{1t}, \dots, \psi_{qt})'$ and retain the q^* largest principal components ψ_t^* except for the largest one. We set $q^* = 3$; (iv) regress $\hat{\varepsilon}_t$ on $1, Y_{t-1}$ and ψ_t^* , and obtain R^2 , the squared uncentred multi-correlation coefficient; (v) compute the test statistic $\text{NN}(p) = TR^2$, and compare it with $\chi_{q^*}^2$.

In addition to the lag order p , Hamilton's (2001) test involves choosing a covariance function $H_p(\cdot)$ for the underlying random field, and the value of the associated $p \times 1$ parameter vector g . Hamilton's (2001) test statistic is computed as follows: (i) set $g_j = 2/(\hat{p}\hat{\sigma}_j^2)^{\frac{1}{2}}$, $j = 1, \dots, p$, where $\hat{\sigma}_j$ is the sample standard deviation of Y_{t-j} ; (ii) calculate the $T \times T$ matrix H whose (t, s) -th element is given by $H_p(\frac{1}{2}[\sum_{j=1}^p g_j^2(Y_{t-j} - Y_{s-j})^2]^{1/2})$, for the pre-specified covariance function $H_p(\cdot)$ given in Hamilton (2001, Theorem 2.2 or Table I); (iii) regress Y_t on 1 and Y_{t-1} , with the estimated residuals $\{\hat{\varepsilon}_t\}$, the residual sample variance $\hat{\sigma}^2 \equiv \hat{\varepsilon}'\hat{\varepsilon}/(T-2)$ and the $T \times T$ projection matrix $M \equiv I_T - X(X'X)^{-1}X$ for X the $T \times 2$ matrix whose t -th row is $(1, Y_{t-1})$; (iv) compute the test statistic

$$\text{HM}(p) = \frac{[\hat{\varepsilon}'H\hat{\varepsilon} - \hat{\sigma}^2\text{tr}(MHM)]^2}{\hat{\sigma}^4\{2\text{tr}[(MHM - (T-2)^{-1}M\text{tr}(MHM))^2]\}},$$

and compare it to the χ_1^2 distribution.

$\hat{f}^{(0,1,0)}(\omega, 0, v)$ in (3.8) and plot the shape of the estimated supremum generalized spectral derivative modulus $m(\omega)$. Here, the Parzen kernel and the data-driven bandwidth \hat{p}_0 described in Section 6 are used.

18. It is not realistic to choose p as large as $p = 20$ (say) for these existing tests when $T = 100$. For example, Tsay's (1986) test would have $\frac{1}{2}p(p+1) = 210$ variables in an auxiliary regression, which is larger than the sample size T . In contrast, the choice of $p = 20$ when $T = 100$ is reasonable for our generalized spectral derivative tests.

19. In fact, the effective sample size in some auxiliary regressions is $T - p$, so we have replaced T with $T - p$ in computing the test statistics of Tsay (1986), White (1989) and Hamilton (2001).

TABLE 1
Empirical levels of tests

α	DGP S.1 : $\varepsilon_t \sim \text{i.i.d. } N(0, 1)$						DGP S.2 : $\varepsilon_t \sim \text{ARCH}(1)$					
	$T = 100$		$T = 250$		$T = 500$		$T = 100$		$T = 250$		$T = 500$	
	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$M_1(2)$	4.4	2.7	5.6	3.3	3.7	2.5	4.4	2.2	6.2	3.2	8.3	5.1
$M_1(4)$	4.9	2.5	7.5	4.3	5.8	3.3	4.1	1.7	5.8	2.8	7.3	4.5
$M_1(6)$	4.5	2.2	7.6	4.0	6.3	3.6	3.6	1.6	5.7	3.2	7.2	3.5
$M_2(2)$	4.6	2.4	5.4	3.3	4.1	2.7	23.7	18.6	27.8	21.0	32.0	23.7
$M_2(4)$	5.6	2.7	7.9	4.3	5.8	3.5	22.2	17.1	27.1	19.9	30.4	22.8
$M_2(6)$	5.7	2.9	8.3	4.7	6.6	4.2	20.3	14.9	24.1	18.8	27.5	21.1
$M_3(2)$	4.7	2.6	5.5	3.3	4.2	2.7	24.8	19.1	28.2	21.3	32.7	24.9
$M_3(4)$	6.3	3.7	7.9	4.7	6.0	3.7	24.0	18.8	27.8	21.2	31.4	24.5
$M_3(6)$	6.5	4.1	8.7	5.5	7.0	4.4	23.0	18.0	25.9	20.3	29.8	23.0
Tsay(1)	10.3	4.6	8.1	3.5	9.1	3.3	24.1	16.7	35.0	27.4	43.3	35.9
Tsay(2)	9.7	5.1	9.0	4.6	7.5	3.9	39.0	30.3	52.4	43.1	61.3	53.6
Tsay(5)	6.7	3.2	7.5	3.6	8.2	4.1	36.6	28.4	57.7	49.3	69.9	62.6
NN(1)	7.2	3.8	9.5	4.5	8.3	3.6	39.3	30.7	57.1	48.4	66.5	59.0
NN(2)	10.7	5.3	9.9	4.7	9.0	4.9	33.5	25.5	43.5	34.7	53.0	45.8
NN(5)	11.1	4.7	11.3	6.5	10.0	5.5	16.0	8.8	20.4	12.6	25.5	16.9
HM(1)	8.5	6.2	6.6	5.5	7.0	4.6	14.5	11.8	18.1	13.5	22.2	18.1
HM(2)	6.8	4.2	6.8	4.0	9.1	5.4	11.6	8.2	13.0	9.7	17.0	12.0
HM(5)	8.4	4.7	9.8	5.8	10.0	5.1	5.7	3.0	7.0	3.9	8.0	4.6

Notes: (1) $\hat{M}_1(\hat{p}_0), \hat{M}_2(\hat{p}_0), \hat{M}_3(\hat{p}_0)$, the generalized spectral tests derived under conditional heteroscedasticity, conditional homoscedasticity and i.i.d., respectively; Tsay(p), Tsay's (1986) test; NN(p), White's (1989) neural network test; HM(p), Hamilton (2001) random field test; (2) DGP S.1, $Y_t = 0.5Y_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{i.i.d. } N(0, 1)$; DGP S.2, $Y_t = 0.5Y_{t-1} + \varepsilon_t, \varepsilon_t = \sqrt{h_t}z_t, z_t \sim \text{i.i.d. } N(0, 1)$; (3) 1000 iterations.

To compute $\hat{M}_a(\hat{p}_0)$, we use the $N(1, 0)$ CDF truncated on $[-3, 3]$ for the weighting function $W(\cdot)$. We use the Bartlett kernel $k_B(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$ for $k(\cdot)$, which has a bounded support and is computationally efficient. Our simulation experience suggests that the choices of $W(\cdot)$ and $k(\cdot)$ have little impact on both the level and the power of our tests.²⁰ 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 $\bar{p} = c[(k_B^{(1)})^2 / \int k_B^2(u)du]^{1/5} T^{1/5} = c(10T)^{1/5}, c > 0$. This rate is optimal for estimating the preliminary generalized spectral derivative $\bar{f}^{(1,1,0)}(\omega, 0, v)$. We choose $c = 2, 4, 6$, which cover a wide range of values for $\bar{p} : [7.96, 23.87]$ for $T = 100, [9.56, 28.69]$ for $T = 250$, and $[10.98, 32.96]$ for $T = 500$.

7.3. Monte Carlo evidence

Table 1 reports the empirical rejection rates of the tests under \mathbb{H}_0 at the 10% and 5% levels, using the asymptotic theory. Under DGP S.1 (homoscedastic errors), all $\hat{M}_a(\hat{p}_0)$ tests underreject \mathbb{H}_0 but not excessively; the tests $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$ derived under conditional homoscedasticity and i.i.d., respectively, have better levels than $\hat{M}_1(\hat{p}_0)$. There is some tendency that a larger preliminary lag order \bar{p} gives a better level for all $\hat{M}_a(p)$ tests. Under DGP S.2 (ARCH errors), $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$ display strong overrejection, as is expected. In contrast, $\hat{M}_1(\hat{p}_0)$ shows

20. We have also used the Parzen kernel (not reported). Although the data-driven lag order \hat{p}_0 is substantially smaller, the test statistics are rather similar to those based on the Bartlett kernel in most cases.

substantial underrejection for $T = 100$, but improves as T increases. Unlike under DGP S.1, a larger \bar{p} does not give a better level for $\hat{M}_1(\hat{\rho}_0)$.

Next, we turn to the levels of TSAY(p), NN(p) and HM(p), which are valid tests under conditional homoscedasticity. Under DGP S.1, these tests have better levels than the $\hat{M}_a(\hat{\rho}_0)$ tests, though not in every case. Under DGP S.2, TSAY(p) and NN(p) show very strong overrejections. The HM(p) test also shows some overrejections when $p = 1, 2$, but interestingly, HM(5) displays underrejection.

Table 2 reports the level-corrected power at the 10% and 5% levels under DGPs P.1–P.8. The empirical critical values are obtained under DGP S.1. We compare $\hat{M}_2(\hat{\rho}_0)$ and $\hat{M}_3(\hat{\rho}_0)$ with TSAY(p), NN(p) and HM(p), all of which are derived under conditional homoscedasticity. Under DGP P.1 (bilinear(1)) and DGP P.2 (nonlinear MA(1)), $\hat{M}_2(\hat{\rho}_0)$ and $\hat{M}_3(\hat{\rho}_0)$ are among the most powerful tests. The power of TSAY(p), NN(p) and HM(p) is sensitive to the choice of p . In contrast, the power of $\hat{M}_2(\hat{\rho}_0)$ and $\hat{M}_3(\hat{\rho}_0)$ is relatively robust to the choice of preliminary lag order \bar{p} . Note that the heteroscedasticity-robust test $\hat{M}_1(\hat{\rho}_0)$ is less powerful than $\hat{M}_2(\hat{\rho}_0)$ and $\hat{M}_3(\hat{\rho}_0)$ under DGPs P.1 and P.2.

Under DGP P.3 (EXP-AR(1)), NN(p) and HM(p) are the most powerful for $p = 1, 2$ and $T = 100$, while TSAY(p) is the least powerful. The $\hat{M}_a(\hat{\rho}_0)$ tests are powerful, though not as powerful as HM(p) and NN(p) with $p = 1, 2$. All $\hat{M}_a(\hat{\rho}_0)$ tests are equally powerful. Under DGP P.4 (SETAR(1)), TSAY(1) and NN(1) are the most powerful, followed by HM(1). The $\hat{M}_a(\hat{\rho}_0)$ tests are equally powerful. They are not as powerful as TSAY(1), NN(1) and HM(1), but they substantially outperform TSAY(5), NN(5) and HM(5). Again, this highlights the robust power property of $\hat{M}_a(\hat{\rho}_0)$ relative to the existing tests.

Under DGP P.5 (STAR(1)), HM(p) is the most powerful, followed by NN(p) and then by TSAY(p). The $\hat{M}_a(\hat{\rho}_0)$ tests are the least powerful against this DGP when $T = 100$, but their power quickly catches up when $T = 250$. The relative ranking of the tests is quite different under DGP P.6 (ARMA(1, 1)). Although it is a first order ARMA process, all $\hat{M}_a(\hat{\rho}_0)$ tests are very powerful and their powers are comparable to the most powerful test NN(2). Moreover, the powers of $\hat{M}_a(\hat{\rho}_0)$ are robust to the choice of the preliminary lag order \bar{p} . In contrast, TSAY(1), NN(1) and HM(1) have no power at all even if T increases. All TSAY(p) tests have little or low power.

The relatively omnibus and robust power performance of the $\hat{M}_a(\hat{\rho}_0)$ tests under DGP P.1–P.6 is encouraging given the fact that DGP P.1–P.6 are all first order dynamic processes whereas $\hat{M}_a(\hat{\rho}_0)$ employs many lags. Such omnibus and robust power apparently comes from the use of the characteristic function and the downward weighting kernel $k(\cdot)$ for lags, which highlights the advantages of the generalized spectrum. In fact, additional merits of $\hat{M}_a(\hat{\rho}_0)$ are further highlighted under DGPs P.7 and P.8. Under DGP P.7 (NMA(5)), which has a declining weight as lag order increases, the $\hat{M}_a(\hat{\rho}_0)$ tests are the most powerful and their powers are robust to the choice of the preliminary lag order \bar{p} . TSAY(p) also has good power and dominates NN(p) and HM(p). Interestingly, although NMA(5) is a fifth order dynamic process, the maximal power of TSAY(p), NN(p) and HM(p) occurs when $p = 2$ instead of $p = 5$.

Under DGP P.8 (SIGN AR(6)), where nonlinearity occurs at lag 6, TSAY(p), NN(p) and HM(p) all have no or low power because $p < 6$. This is the case even when T increases. Thus, it is important to select a suitable lag order for these tests, but this requires knowledge of the alternative. In contrast, the $\hat{M}_a(\hat{\rho}_0)$ tests are very powerful, and they require no knowledge of the lag structure in the DGP!

In summary, we observe:

- (1) The empirical levels of the $\hat{M}_a(\hat{\rho}_0)$ tests are smaller than the nominal levels, but they improve as the sample size increases. Under homoscedastic errors, the homoscedasticity-

TABLE 2
Empirical powers of tests

α	DGP P.1.: Bilinear(1)						DGP P.2.: NMA(1)						DGP P.3.: EXP-AR(1)						DGP P.4.: SETAR(1)					
	T = 100			T = 250			T = 100			T = 250			T = 100			T = 250			T = 100			T = 250		
	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%		
$\hat{M}_1(2)$	55.8	42.8	74.8	64.0	78.4	62.4	99.4	99.9	86.0	82.2	100.0	99.0	58.2	46.6	92.4	86.2								
$\hat{M}_1(4)$	48.2	35.0	64.8	52.4	68.4	51.8	98.4	96.0	86.4	80.4	99.8	98.6	49.4	37.6	86.0	76.2								
$\hat{M}_1(6)$	43.2	29.8	60.2	44.8	64.4	44.8	97.2	92.6	84.6	77.4	99.4	98.8	45.2	31.2	79.2	66.4								
$\hat{M}_2(2)$	93.8	87.2	99.8	99.8	89.8	82.0	99.8	99.4	84.2	80.0	99.4	99.0	59.4	48.0	93.4	88.0								
$\hat{M}_2(4)$	89.4	78.0	99.4	98.2	83.2	65.6	99.8	99.2	85.6	77.2	99.6	98.6	51.8	37.0	86.8	77.8								
$\hat{M}_2(6)$	84.6	70.8	98.0	95.6	75.4	56.8	99.8	98.2	81.8	74.8	99.2	98.4	45.2	28.2	79.6	68.0								
$\hat{M}_3(2)$	94.4	87.8	99.8	99.8	90.0	82.2	99.8	99.4	84.4	80.2	99.6	99.0	59.2	48.0	93.6	88.0								
$\hat{M}_3(4)$	90.4	81.2	99.8	98.4	83.2	66.6	99.8	99.2	85.6	77.8	99.6	98.6	52.2	37.4	86.8	78.2								
$\hat{M}_3(6)$	85.6	72.8	98.8	96.2	75.6	57.8	99.8	98.4	82.4	75.4	99.4	98.4	45.8	28.4	79.8	67.8								
Tsay(1)	35.4	27.8	48.0	41.0	43.4	34.6	62.0	54.6	36.8	28.2	39.0	32.4	80.2	70.6	99.8	98.2								
Tsay(2)	86.4	81.2	94.2	92.2	74.0	62.8	93.2	88.8	23.8	17.8	29.2	21.2	67.4	53.4	97.0	94.0								
Tsay(5)	84.8	77.6	96.6	94.2	57.8	42.2	88.0	80.8	15.4	11.6	18.6	12.2	38.4	25.4	85.0	75.4								
NN(1)	91.2	87.2	94.8	93.6	95.6	91.6	99.8	99.8	98.2	95.6	100.0	100.0	78.2	68.0	98.2	96.8								
NN(2)	72.0	63.2	90.4	85.2	50.0	41.4	70.2	64.0	90.6	85.8	100.0	100.0	52.6	39.8	82.0	76.2								
NN(5)	26.2	18.8	48.0	36.4	14.2	9.0	17.6	8.8	78.4	70.8	98.4	97.0	10.4	6.4	11.6	5.2								
HM(1)	85.8	74.0	99.6	98.0	88.2	79.0	100.0	99.4	100.0	100.0	100.0	100.0	66.2	53.4	96.6	93.8								
HM(2)	92.0	85.4	100.0	100.0	92.0	86.0	100.0	100.0	100.0	100.0	100.0	100.0	46.8	38.0	87.2	83.4								
HM(5)	38.6	27.8	86.6	73.4	33.8	23.0	91.2	82.4	95.6	92.8	100.0	100.0	17.2	12.0	41.6	29.8								

TABLE 2—continued

α	DGP P.5: STAR(1)			DGP P.6: ARMA(1,1)			DGP P.7: NMA(5)			DGP P.8: SIGN AR(6)						
	T = 100		T = 250	T = 100		T = 250	T = 100		T = 250	T = 100		T = 250				
	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%				
$\hat{M}_1(2)$	58.6	45.0	96.6	92.8	86.8	76.0	99.8	99.6	96.0	90.2	100.0	100.0	52.8	43.6	100.0	100.0
$\hat{M}_1(4)$	54.6	38.0	95.4	89.0	83.6	70.4	99.6	99.0	90.0	85.0	100.0	100.0	99.2	97.2	100.0	100.0
$\hat{M}_1(6)$	52.0	34.6	92.6	80.6	78.8	65.8	99.4	98.0	87.2	78.6	100.0	100.0	99.8	99.2	100.0	100.0
$\hat{M}_2(2)$	61.2	48.0	96.8	93.4	87.2	79.6	99.8	99.8	97.6	94.0	100.0	100.0	53.6	42.6	100.0	100.0
$\hat{M}_2(4)$	60.2	39.0	95.6	90.4	85.4	70.8	99.6	99.0	94.0	87.2	100.0	100.0	99.4	96.8	100.0	100.0
$\hat{M}_2(6)$	51.8	33.0	93.2	83.0	79.6	66.4	99.4	98.2	91.0	81.4	100.0	100.0	100.0	99.2	100.0	100.0
$\hat{M}_3(2)$	61.0	47.6	96.8	93.4	87.4	79.2	99.8	99.8	97.6	93.6	100.0	100.0	53.6	42.8	100.0	100.0
$\hat{M}_3(4)$	60.2	39.0	95.4	90.4	85.4	71.0	99.6	99.0	94.0	87.6	100.0	100.0	99.4	97.0	100.0	100.0
$\hat{M}_3(6)$	52.4	33.6	93.2	83.0	79.8	66.6	99.4	98.2	91.4	81.6	100.0	100.0	100.0	99.2	100.0	100.0
Tsay(1)	90.4	85.8	99.6	99.6	2.8	0.8	4.0	0.6	78.8	65.8	99.0	98.2	9.2	5.8	12.6	7.2
Tsay(2)	87.0	77.0	100.0	99.4	13.0	6.2	12.6	7.0	97.4	95.0	100.0	100.0	9.6	5.2	14.2	7.4
Tsay(5)	73.4	62.2	100.0	99.8	13.8	8.8	17.0	11.4	86.0	77.2	100.0	100.0	33.2	22.8	40.0	31.4
NN(1)	100.0	100.0	100.0	100.0	3.0	1.2	2.0	0.6	79.4	72.8	99.0	98.2	12.6	5.0	11.6	7.8
NN(2)	94.0	91.2	100.0	99.4	90.0	83.4	100.0	99.8	69.0	60.8	88.0	83.2	19.0	11.2	26.6	17.6
NN(5)	83.6	74.6	99.4	98.0	78.4	66.8	95.8	93.0	20.0	12.4	20.6	14.8	43.0	33.4	47.2	38.2
HM(1)	100.0	100.0	100.0	100.0	4.0	1.0	2.2	0.4	64.4	50.8	97.0	93.6	8.8	3.8	12.2	6.0
HM(2)	100.0	100.0	100.0	100.0	51.6	43.4	94.6	91.6	82.0	78.0	100.0	100.0	20.8	15.6	20.6	17.0
HM(5)	100.0	100.0	100.0	100.0	24.0	15.8	63.8	48.6	49.0	38.2	94.4	89.8	44.2	35.4	50.8	40.4

Notes: (1) DGP P.1, $Y_t = 0.5Y_{t-1} + 0.6Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$; DGP P.2, $Y_t = 0.5Y_{t-1} - 0.6\varepsilon_{t-1}^2 + \varepsilon_t$; DGP P.3, $Y_t = 0.5Y_{t-1} + 10Y_{t-1} \exp(-Y_{t-1}^2) + \varepsilon_t$; DGP P.4, $Y_t = 0.5Y_{t-1}\mathbf{1}(Y_{t-1} \leq 0) - 0.5Y_{t-1}\mathbf{1}(Y_{t-1} > 0) + \varepsilon_t$; DGP P.5, $Y_t = 1 - 0.5Y_{t-1} + (-4 - 0.4Y_{t-1})G(\alpha Y_{t-1}) + \varepsilon_t$, where $G(z) = [1 + \exp(-z)]^{-1}$; DGP P.6, $Y_t = 0.5Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$; DGP P.7, $Y_t = 0.5Y_{t-1} + \sum_{j=1}^5 0.5^j \varepsilon_{t-j}^2 + \varepsilon_t$; DGP P.8, $Y_t = \mathbf{1}(Y_{t-6} > 0) - \mathbf{1}(Y_{t-6} < 0) + \varepsilon_t$, where $\varepsilon_t \sim$ i.i.d. $N(0, 1)$; (2) 500 iterations.

specific tests, $\hat{M}_2(\hat{p}_0)$ and $\hat{M}_3(\hat{p}_0)$, have better levels than the heteroscedasticity-robust test $\hat{M}_1(\hat{p}_0)$. Under ARCH errors, $\hat{M}_1(\hat{p}_0)$ continues to have reasonable levels (though still underrejected), but all homoscedasticity-specific tests strongly overreject the correct model.

- (2) The powers of $TSAY(p)$, $NN(p)$ and $HM(p)$ are rather sensitive to the choice of the lag order p , and the maximal powers of these tests usually do not arise when p matches the true lag order of the DGP. It is often beneficial to use more lags than those specified in the null model, but there is no rule to select an optimal lag order for these tests. In contrast, the $\hat{M}_a(\hat{p}_0)$ tests have relatively robust power with respect to the choice of the preliminary lag order \bar{p} , and they require no knowledge of the lag structure of the potential alternative.
- (3) The $\hat{M}_a(\hat{p}_0)$ tests are not always the most powerful in detecting each of the eight DGPs. However, they have relatively omnibus power against all eight DGPs provided the sample size is sufficiently large. $TSAY(p)$, $NN(p)$ and $HM(p)$ can be very powerful in detecting some DGPs but may have little power against others even when the sample size increases.
- (4) The heteroscedasticity-robust generalized spectral test has similar power to the homoscedasticity-specific generalized spectral tests in most cases. However, there exist cases (bilinear (1) and nonlinear MA(1)) where the former is less powerful than the latter tests.

8. PREDICTABILITY AND NONLINEARITY OF STOCK RETURNS

We now use our tests to check the predictability of stock returns. Although several seemingly anomalous departures from market efficiency have been well documented, many financial economists still believe that no other proposition in economics has more solid empirical support than the efficient market hypothesis (EMH). The majority of studies on EMH has focused on the predictability of common stock returns. Using the variance ratio test, which is robust to conditional heteroscedasticity,²¹ Lo and MacKinlay (1988) found significant positive serial correlation for weekly and monthly stock returns.

It has been argued in the literature that significant autocorrelation in stock returns may be superficial: because small capitalization stocks trade less frequently than large stocks, new information is impounded first into large capitalization stock prices and then into small capitalization stock prices with a time lag. This time lag will induce a positive serial correlation in the weighted index of stock returns. Thus, the rejection of EMH for stock prices using autocorrelation may be the result of this nonsynchronous phenomenon. Using a hypothesized model, Lo and MacKinlay (1988) showed that nonsynchronous trading can at most account for part of the documented positive autocorrelation and concluded that stock returns are predictable. However, this model has not been empirically tested. Furthermore, Boudoukh, Richardson and Whitelaw (1994) showed that inferences from nonsynchronous trading models are highly sensitive to the assumptions on nontrading intervals and other stock market parameters, and they argued that prior studies understated the effects of nonsynchronous trading. It seems difficult to quantify how much of the significant autocorrelation could be attributed to nonsynchronous trading.

Here, we further contribute to this literature by taking a new approach. Instead of arguing how much the significant autocorrelation can be explained by nonsynchronous trading, we check whether stock returns are still predictable after removing all linear serial correlation. In other words, we check whether there exist predictable nonlinear components in mean. Lo and

21. The variance-ratio test can be interpreted as a test based on the power spectral density estimator at frequency zero, using the Bartlett kernel. See, *e.g.* Cochrane (1988).

TABLE 3
Predictability and normality of stock returns

$\hat{M}_1(\hat{\rho}_0)$ c	Martingale testing				Linearity testing			
	S&P 500		NASDAQ		S&P 500		NASDAQ	
	Statistics	p -value	Statistics	p -value	Statistics	p -value	Statistics	p -value
1	13.85	0.000	30.04	0.000	2.57	0.005	7.78	0.000
2	13.85	0.000	27.19	0.000	2.57	0.005	7.04	0.000
3	12.98	0.000	25.11	0.000	2.47	0.007	6.50	0.000
4	12.27	0.000	23.51	0.000	2.37	0.009	6.11	0.000
5	11.70	0.000	22.19	0.000	2.28	0.011	5.80	0.000
6	11.25	0.000	21.10	0.000	2.23	0.013	5.55	0.000
7	10.90	0.000	20.23	0.000	2.18	0.014	5.36	0.000
8	10.61	0.000	19.51	0.000	2.15	0.016	5.21	0.000
9	14.65	0.000	32.43	0.000	2.69	0.004	8.40	0.000
10	14.65	0.000	32.43	0.000	2.69	0.004	8.40	0.000

Notes: (1) The sample period for both S&P 500 and NASDAQ daily returns is from 1 December 1972 to 31 December 2001; (2) Martingale testing is the application of $\hat{M}_1(\hat{\rho})$ to the raw stock returns $Y_t = 100 \ln(P_t/P_{t-1})$ where P_t is the stock price at day t ; (3) Linearity testing is the application of $\hat{M}_1(\hat{\rho})$ to the estimated residuals of a linear AR(d) model with intercept, where the order d is selected by the BIC criterion. For S&P 500, the selected linear model is AR(2); for NASDAQ, the selected linear model is AR(1); (4) The data driven lag order $\hat{\rho}_0$ is computed using formula (6.4), with the preliminary bandwidth $\bar{p} = c(10T)^{1/5}$, where c ranges from 1 to 10.

MacKinlay's (1988) variance ratio test cannot be used for this purpose, because it can only capture linear dependence.

We consider two daily stock price indices: S&P 500 index and NASDAQ index, from 1 December 1972 to 31 December 2001, obtained from CRSP. Define stock return $Y_t = 100 \ln(P_t/P_{t-1})$, where P_t is the closing stock price index at day t . A graphical examination shows strong volatility clustering for both returns series. Thus the heteroscedasticity-robust test $\hat{M}_1(\hat{\rho}_0)$ should be used.

We first use $\hat{M}_1(\hat{\rho}_0)$ to test whether the stock return series $\{Y_t\}$ is a *m.d.s.* Table 3, under subtitle "martingale testing", reports the $\hat{M}_1(\hat{\rho}_0)$ statistics using the data-driven lag order $\hat{\rho}_0$ in (6.4), with the preliminary bandwidth $\bar{p} = c(10T)^{1/5}$, for $c = 1, \dots, 10$. These statistics are quite robust to the choice of \bar{p} and have essentially zero asymptotic p -values, suggesting strong evidence against EMH for both S&P 500 and NASDAQ daily returns. Figure 1(a, b) displays the shapes of the supremum generalized spectral derivative modulus $m(\omega)$ in (3.5) for both S&P 500 and NASDAQ returns. They are apparently nonuniform over the frequency ω , confirming the nonmartingale behaviour of stock prices. In particular, there exists more persistent serial dependence in mean for NASDAQ daily returns than for S&P 500 returns, because NASDAQ returns have a sharp mode in $m(\omega)$ at frequency zero.

Next, we examine whether stock returns contain predictable nonlinearities in mean after removing linear dependence. This checks if stock returns are linear in mean. We first estimate an AR(d) model

$$Y_t = \alpha_0 + \sum_{j=1}^d \alpha_j Y_{t-j} + \varepsilon_t,$$

where lag order d is selected via the BIC criterion, which delivers a consistent order selection for stationary linear processes (Hannan, 1980). The selected models are an AR(2) for S&P 500, and an AR(1) for NASDAQ, with small but significant AR coefficients.²² Table 3, under the

22. Alternatively we have also considered a class of ARMA(p, q) models with intercept, where the order (p, q) is selected by the BIC criterion. An MA(1) model with intercept is selected for both S&P 500 and NASDAQ daily returns.

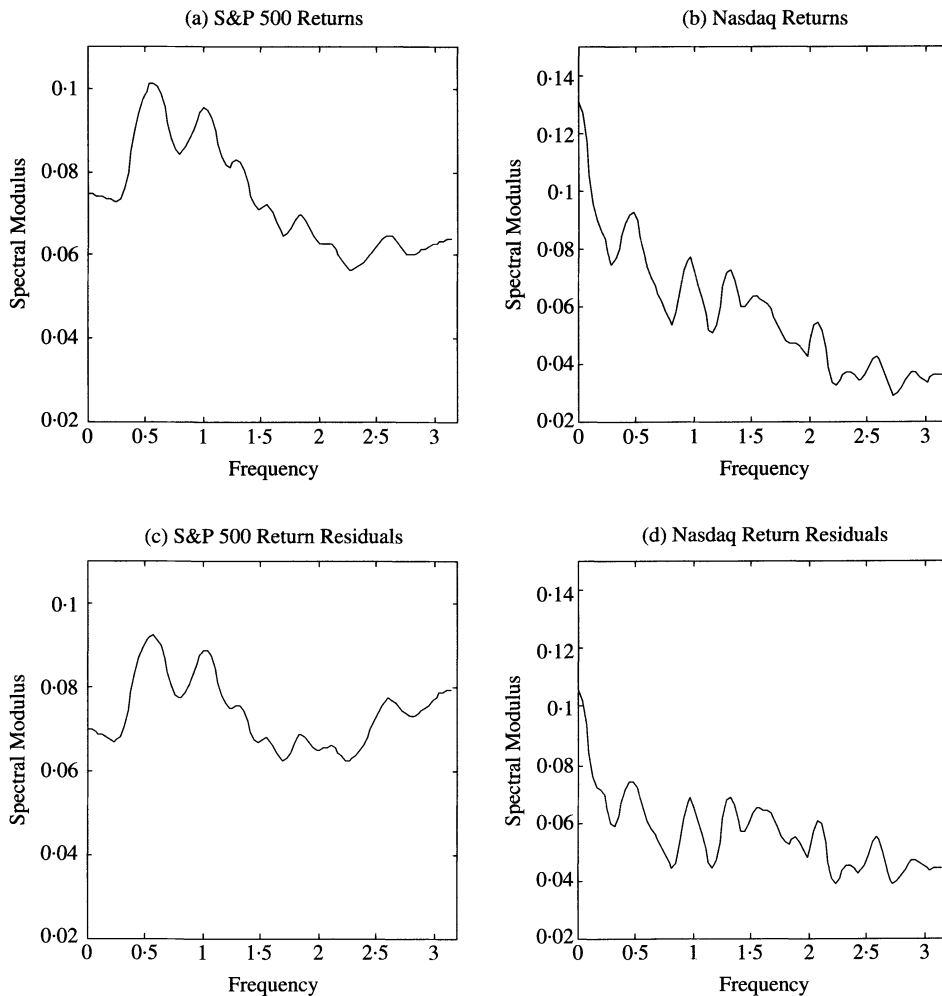


FIGURE 1

Supremum generalized spectral derivative modulus $m(\omega)$

subtitle “linearity testing”, reports the $\hat{M}_1(\hat{\rho}_0)$ statistics applied to the OLS $AR(d)$ residuals. These statistic values are much smaller than the $\hat{M}_1(\hat{\rho}_0)$ statistics applied to the raw return data, but they are still significant. The asymptotic p -values are below 2 for S&P 500 and are essentially zero for NASDAQ. These results suggest that stock returns contain predictable nonlinearities in mean, and the evidence is stronger for NASDAQ than S&P 500. Again, Figure 1(c, d) displays the shapes of the supremum generalized spectral derivative modulus $m(\omega)$ for two residual series, which are still nonuniform over ω . Still, NASDAQ return residuals have a sharp mode for $m(\omega)$ at frequency zero, indicating that there exists more persistent nonlinear dependence in mean for NASDAQ daily returns than for S&P 500 daily returns. Interestingly, Figure 1 shows that there are two mild spectral modes at around $\omega = 0.6$ and 1.2 for both stock returns and their residuals. These correspond to some mild cyclical dynamics with periodicities of about 12 days and 6 days, respectively, which might be due to the well-known “calendar effects”.

The $\hat{M}_1(\hat{\rho}_0)$ test statistics obtained are very close to those under the $AR(d)$ specification.

9. CONCLUSION

Using a generalized spectral derivative approach, we develop a class of residual-based, generally applicable specification tests for linear and nonlinear conditional mean models in time series, where the dimension of the conditioning information set may be infinite. The tests can detect a wide range of model misspecification in mean while being robust to conditional heteroscedasticity and other higher order time-varying moments of unknown form. They check a large number of lags but naturally discount higher order lags, which alleviates the power loss due to the loss of a large number of degrees of freedom. The tests enjoy the appealing “nuisance parameter free” property that parameter estimation uncertainty has no impact on the limit distribution of the tests. A simulation study shows that it is important to take into account the impact of conditional heteroscedasticity to ensure a proper level. The tests have omnibus and robust power against a variety of dynamic misspecification and nonlinear alternatives in mean relative to some existing tests. We use our tests to check the predictability of stock price changes. After removing significant but possibly spurious autocorrelations due to nonsynchronous trading, we still find significant nonlinearity in mean for S&P 500 and NASDAQ daily returns.

MATHEMATICAL APPENDIX

Throughout the Appendix, we let $M_a(p)$, $a = 1, 2, 3$, be defined in the same way as $\hat{M}_a(p)$ in (3.11)–(3.13), with the unobservable sample $\{\varepsilon_t \equiv \varepsilon_t(\theta_0)\}_{t=1}^T$, where $\theta_0 \equiv p \lim \hat{\theta}$, replacing the estimated residual sample $\{\hat{\varepsilon}_t\}_{t=1}^T$ defined in (3.7). Also, $C \in (1, \infty)$ denotes a generic bounded constant.

Proof of Theorem 1. For space, we only consider $\hat{M}_1(p)$; the proofs for $\hat{M}_2(p)$ and $\hat{M}_3(p)$ are similar and a bit simpler. It suffices to show Theorems A1–A3 below. Theorem A1 implies that the use of $\{\hat{\varepsilon}_t\}_{t=1}^T$ rather than $\{\varepsilon_t\}_{t=1}^T$ has no impact on the limit distribution of $\hat{M}_1(p)$. Theorem A2 implies that the use of the truncated disturbances $\{\varepsilon_{q,t}\}_{t=1}^T$ rather than $\{\varepsilon_t\}_{t=1}^T$ has no impact on the limit distribution of $\hat{M}_1(p)$ for q sufficiently large. The assumption that $\varepsilon_{q,t}$ is independent of $\{\varepsilon_{t-j}\}_{j=q+1}^\infty$ when q is large simplifies a great deal the proof of asymptotic normality of $\hat{M}_1(p)$.

Theorem A1. Under the conditions of Theorem 1, $\hat{M}_1(p) - M_1(p) \xrightarrow{P} 0$.

Theorem A2. Let $M_{1q}(p)$ be defined as $M_1(p)$ with $\{\varepsilon_{q,t}\}_{t=1}^T$ replacing $\{\varepsilon_t\}_{t=1}^T$, where $\{\varepsilon_{q,t}\}$ is as in Assumption A2. Then under the conditions of Theorem 1 and $q = p^{1+\frac{1}{4b-2}} (\ln^2 T)^{\frac{1}{2b-1}}$, $M_{1q}(p) - M_1(p) \xrightarrow{P} 0$.

Theorem A3. Under the conditions of Theorem 1 and $q = p^{1+\frac{1}{4b-2}} (\ln^2 T)^{\frac{1}{2b-1}}$, $M_{1q}(p) \xrightarrow{d} N(0, 1)$.

Proof of Theorem A1. Noting that $\varepsilon_t(\theta) \equiv Y_t - g(I_{t-1}, \theta)$ in (3.1), where I_{t-1} is the unobservable information set from period t to the infinite past, we write $\hat{\varepsilon}_t \equiv Y_t - g(I_{t-1}^\dagger, \hat{\theta}) = \varepsilon_t(\hat{\theta}) + g(I_{t-1}, \hat{\theta}) - g(I_{t-1}^\dagger, \hat{\theta})$. Note that $\hat{\varepsilon}_t \neq \varepsilon_t(\hat{\theta})$ because $I_{t-1}^\dagger \neq I_{t-1}$ generally, and Assumption A5 implies

$$\sum_{t=1}^T [\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})]^2 = \sum_{t=1}^T |g(I_{t-1}, \hat{\theta}) - g(I_{t-1}^\dagger, \hat{\theta})|^2 = O_P(1). \quad (\text{A.1})$$

By the mean value theorem, we have $\varepsilon_t(\hat{\theta}) = \varepsilon_t - g'_t(\bar{\theta})'(\hat{\theta} - \theta_0)$ for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0 , where $g'_t(\theta) \equiv \frac{\partial}{\partial \theta} g(I_{t-1}, \theta)$. It follows from the Cauchy–Schwarz inequality, and Assumptions A3 and A4 that

$$\sum_{t=1}^T [\varepsilon_t(\hat{\theta}) - \varepsilon_t]^2 \leq T \|\hat{\theta} - \theta_0\|^2 T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_0} \|g'_t(\theta)\|^2 = O_P(1), \quad (\text{A.2})$$

where Θ_0 is a neighbourhood of θ_0 . Both (A.1) and (A.2) imply

$$\sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 = O_P(1). \quad (\text{A.3})$$

Now, put $T_j \equiv T - |j|$, and let $\tilde{\sigma}_j^{(1,0)}(0, v)$ be defined in the same way as $\hat{\sigma}_j^{(1,0)}(0, v)$ in (3.8), with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. To show $\hat{M}_1(p) - M_1(p) \xrightarrow{P} 0$, it suffices to show

$$\hat{D}_1^{-\frac{1}{2}}(p) \int \sum_{j=1}^{T-1} k^2(j/p) T_j [|\hat{\sigma}_j^{(1,0)}(0, v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0, v)|^2] dW(v) \xrightarrow{P} 0, \tag{A.4}$$

$\hat{C}_1(p) - \tilde{C}_1(p) = O_P(T^{-\frac{1}{2}})$, and $\hat{D}_1(p) - \tilde{D}_1(p) \xrightarrow{P} 0$, where $\tilde{C}_1(p)$ and $\tilde{D}_1(p)$ are defined in the same way as $\hat{C}_1(p)$ and $\hat{D}_1(p)$ in (3.11), with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. For space, we focus on the proof of (A.4); the proofs for $\hat{C}_1(p) - \tilde{C}_1(p) = O_P(T^{-\frac{1}{2}})$ and $\hat{D}_1(p) - \tilde{D}_1(p) \xrightarrow{P} 0$ are straightforward. We note that it is necessary to obtain the convergence rate $O_P(T^{-\frac{1}{2}})$ for $\hat{C}_1(p) - \tilde{C}_1(p)$ to ensure that replacing $\hat{C}_1(p)$ with $\tilde{C}_1(p)$ has asymptotically negligible impact given $p/T \rightarrow 0$.

To show (A.4), we first decompose

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j [|\hat{\sigma}_j^{(1,0)}(0, v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0, v)|^2] dW(v) = \hat{A}_1 + 2\text{Re}(\hat{A}_2), \tag{A.5}$$

where

$$\begin{aligned} \hat{A}_1 &= \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left| \hat{\sigma}_j^{(1,0)}(0, v) - \tilde{\sigma}_j^{(1,0)}(0, v) \right|^2 dW(v), \\ \hat{A}_2 &= \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left[\hat{\sigma}_j^{(1,0)}(0, v) - \tilde{\sigma}_j^{(1,0)}(0, v) \right] \tilde{\sigma}_j^{(1,0)}(0, v)^* dW(v), \end{aligned}$$

where $\text{Re}(\hat{A}_2)$ is the real part of \hat{A}_2 and $\tilde{\sigma}_j^{(1,0)}(0, v)^*$ is the complex conjugate of $\tilde{\sigma}_j^{(1,0)}(0, v)$. Then, (A.4) follows from Propositions A1 and A2 below, and $p \rightarrow \infty$ as $T \rightarrow \infty$.

Proposition A1. Under the conditions of Theorem 1, $\hat{A}_1 = O_P(1)$.

Proposition A2. Under the conditions of Theorem 1, $p^{-\frac{1}{2}} \hat{A}_2 \xrightarrow{P} 0$.

Proof of Proposition A1. Put $\hat{\delta}_t(v) \equiv e^{iv\hat{\varepsilon}_t} - e^{iv\varepsilon_t}$ and $\psi_t(v) \equiv e^{iv\varepsilon_t} - \varphi(v)$, where, as before, $\varphi(v) \equiv E(e^{iv\varepsilon_t})$. Then straightforward algebra yields that for $j > 0$,

$$\begin{aligned} \hat{\sigma}_j^{(1,0)}(0, v) - \tilde{\sigma}_j^{(1,0)}(0, v) &= iT_j^{-1} \sum_{t=j+1}^T (\hat{\varepsilon}_t - \varepsilon_t) \hat{\delta}_{t-j}(v) - i \left[T_j^{-1} \sum_{t=j+1}^T (\hat{\varepsilon}_t - \varepsilon_t) \right] \\ &\quad \times \left[T_j^{-1} \sum_{t=j+1}^T \hat{\delta}_{t-j}(v) \right] + iT_j^{-1} \sum_{t=j+1}^T \varepsilon_t \hat{\delta}_{t-j}(v) \\ &\quad - i \left(T_j^{-1} \sum_{t=j+1}^T \varepsilon_t \right) \left[T_j^{-1} \sum_{t=j+1}^T \hat{\delta}_{t-j}(v) \right] \\ &\quad + iT_j^{-1} \sum_{t=j+1}^T (\hat{\varepsilon}_t - \varepsilon_t) \psi_{t-j}(v) \\ &\quad - i \left[T_j^{-1} \sum_{t=j+1}^T (\hat{\varepsilon}_t - \varepsilon_t) \right] \left[T_j^{-1} \sum_{t=j+1}^T \psi_{t-j}(v) \right] \\ &= i[\hat{B}_{1j}(v) - \hat{B}_{2j}(v) + \hat{B}_{3j}(v) - \hat{B}_{4j}(v) + \hat{B}_{5j}(v) - \hat{B}_{6j}(v)], \text{ say.} \end{aligned} \tag{A.6}$$

It follows that $\hat{A}_1 \leq 8 \sum_{a=1}^6 \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{aj}(v)|^2 dW(v)$. Proposition A1 follows from Lemmas A1 to A6 below, and $p/T \rightarrow 0$.

Lemma A1. $\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{1j}(v)|^2 dW(v) = O_P(p/T)$.

Lemma A2. $\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{2j}(v)|^2 dW(v) = O_P(p/T)$.

Lemma A3. $\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{3j}(v)|^2 dW(v) = O_P(p/T)$.

Lemma A4. $\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{4j}(v)|^2 dW(v) = O_P(p/T)$.

Lemma A5. $\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{5j}(v)|^2 dW(v) = O_P(1)$.

Lemma A6. $\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{B}_{6j}(v)|^2 dW(v) = O_P(p/T)$.

We now show these lemmas. Throughout, we put $a_T(j) \equiv k^2(j/p)T_j^{-1}$.

Proof of Lemma A1. By the Cauchy–Schwarz inequality and the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued variables z_1 and z_2 , we have

$$|\hat{B}_{1j}(v)|^2 \leq \left[T_j^{-1} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 \right] \left[T_j^{-1} \sum_{t=1}^T |\hat{\delta}_t(v)|^2 \right] \leq v^2 \left[T_j^{-1} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 \right]^2.$$

It follows from (A.3), and Assumptions A6 and A7 that

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

where we made use of the fact that

$$\sum_{j=1}^{T-1} a_T(j) = \sum_{j=1}^{T-1} k^2(j/p)T_j^{-1} = O(p/T) \tag{A.7}$$

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

Proof of Lemma A2. By the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued z_1 and z_2 , we have

$$|\hat{B}_{2j}(v)|^2 \leq \left[T_j^{-1} \sum_{t=1}^T |\hat{\varepsilon}_t - \varepsilon_t| \right]^2 \left[T_j^{-1} \sum_{t=1}^T |v\hat{\varepsilon}_t - v\varepsilon_t| \right]^2 \leq v^2 \left[T_j^{-1} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 \right]^2.$$

The desired result then follows using reasoning similar to that of Lemma A1. \parallel

Proof of Lemma A3. We decompose

$$\hat{B}_{3j}(v) = T_j^{-1} \sum_{t=j+1}^T \varepsilon_t [e^{iv\hat{\varepsilon}_{t-j}} - e^{iv\varepsilon_{t-j}(\hat{\theta})}] + T_j^{-1} \sum_{t=j+1}^T \varepsilon_t [e^{iv\varepsilon_{t-j}(\hat{\theta})} - e^{iv\varepsilon_{t-j}}] \equiv \hat{B}_{31j}(v) + \hat{B}_{32j}(v). \tag{A.8}$$

First, we consider $\hat{B}_{31j}(v)$. Using the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued z_1 and z_2 , Minkowski’s inequality, and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} E|\hat{B}_{31j}(v)|^2 &\leq v^2 E \left[T_j^{-1} \sum_{t=j+1}^T |\varepsilon_t| |\hat{\varepsilon}_{t-j} - \varepsilon_{t-j}(\hat{\theta})| \right]^2 \\ &\leq v^2 \left\{ T_j^{-1} \sum_{t=j+1}^T (E\varepsilon_t^4)^{\frac{1}{4}} [E(\sup_{\theta \in \Theta_0} |g(I_{t-j}^\dagger, \theta) - g(I_{t-j}, \theta)|)^4]^{\frac{1}{4}} \right\}^2 \\ &\leq CT_j^{-2} v^2 \left\{ \sum_{t=1}^T [E(\sup_{\theta \in \Theta_0} |g(I_t^\dagger, \theta) - g(I_t, \theta)|)^4]^{\frac{1}{4}} \right\}^2. \end{aligned}$$

It follows from Markov’s inequality, (A.7), and Assumptions A1, A5 and A7 that

$$\sum_{j=1}^{T-1} k^2(j/p)T_j |\hat{B}_{31j}(v)|^2 = O_P(p/T). \tag{A.9}$$

Next, we consider $\hat{B}_{32j}(v)$. Using the inequality that $|e^{iz} - 1 - iz| \leq |z|^2$ for any real-valued z , we have

$$|e^{iv\varepsilon_{t-j}(\hat{\theta})} - e^{iv\varepsilon_{t-j}} - iv[\varepsilon_{t-j}(\hat{\theta}) - \varepsilon_{t-j}]e^{iv\varepsilon_{t-j}}| \leq v^2 [\varepsilon_{t-j}(\hat{\theta}) - \varepsilon_{t-j}]^2. \tag{A.10}$$

Also, a second order Taylor series expansion yields

$$\varepsilon_{t-j}(\hat{\theta}) = \varepsilon_{t-j} - g'_{t-j}(\theta_0)'(\hat{\theta} - \theta_0) - \frac{1}{2}(\hat{\theta} - \theta_0)' g''_{t-j}(\bar{\theta})(\hat{\theta} - \theta_0) \tag{A.11}$$

for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0 where $g''_{t-j}(\theta) \equiv \frac{\partial^2}{\partial \theta \partial \theta'} g(I_{t-j}, \theta)$. Put $\xi_t(v) = g'_t(\theta_0)e^{iv\varepsilon_t}$. Then (A.10) and (A.11) imply

$$|e^{iv\varepsilon_{t-j}(\hat{\theta})} - e^{iv\varepsilon_{t-j}} - iv\xi_{t-j}(v)(\hat{\theta} - \theta_0)| \leq v^2 [\varepsilon_{t-j}(\hat{\theta}) - \varepsilon_{t-j}]^2 + |v| \|\hat{\theta} - \theta_0\|^2 \sup_{\theta \in \Theta_0} \|g''_{t-j}(\theta)\|.$$

Therefore, by the definition of $\hat{B}_{32j}(v)$, we obtain

$$T_j |\hat{B}_{32j}(v)| \leq |v| \|\hat{\theta} - \theta_0\| \left\| \sum_{t=j+1}^T \varepsilon_t \xi_{t-j}(v) \right\| + v^2 \sum_{t=j+1}^T |\varepsilon_t| [\varepsilon_{t-j}(\hat{\theta}) - \varepsilon_{t-j}]^2 + |v| \|\hat{\theta} - \theta_0\|^2 \sum_{t=j+1}^T |\varepsilon_t| \sup_{\theta \in \Theta_0} \|g''_{t-j}(\theta)\|.$$

It follows from Assumptions A1 to A8 and (A.7) that

$$\begin{aligned} \sum_{j=1}^{T-1} \int k^2(j/p) T_j |\hat{B}_{32j}(v)|^2 dW(v) &\leq 4\|\sqrt{T}(\hat{\theta} - \theta_0)\|^2 \sum_{j=1}^{T-1} k^2(j/p) \\ &\times \int \left\| T_j^{-1} \sum_{t=j+1}^T \varepsilon_t \xi_{t-j}(v) \right\|^2 v^2 dW(v) + 4\|\sqrt{T}(\hat{\theta} - \theta_0)\|^4 \\ &\times \left(T^{-1} \sum_{t=1}^T \varepsilon_t^2 \right) \left\{ T^{-1} \sum_{t=1}^T [\sup_{\theta \in \Theta_0} \|g'_t(\theta)\|]^4 \right\} \\ &\times \left[\sum_{j=1}^{T-1} a_T(j) \right] \int v^4 dW(v) + 4\|\sqrt{T}(\hat{\theta} - \theta_0)\|^4 \left(T^{-1} \sum_{t=1}^T \varepsilon_t^2 \right) \\ &\times \left\{ T^{-1} \sum_{t=1}^T [\sup_{\theta \in \Theta_0} \|g''_t(\theta)\|]^2 \right\} \left[\sum_{j=1}^{T-1} a_T(j) \right] \int v^2 dW(v) \\ &= O_P(p/T), \end{aligned} \tag{A.12}$$

where we made use of the fact that $E \left\| \sum_{t=j+1}^T \varepsilon_t \xi_{t-j}(v) \right\|^2 \leq CT_j$ given $E(\varepsilon_t | I_{t-1}) = 0$ a.s. under \mathbb{H}_0 and Assumptions A1 and A3. The desired result of Lemma A5 follows from (A.8), (A.9) and (A.12). \parallel

Proof of Lemma A4. By the Cauchy–Schwarz inequality, $|\hat{B}_{4j}(v)|^2 \leq (T_j^{-1} \sum_{t=j+1}^T \varepsilon_t)^2 T_j^{-1} \sum_{t=j+1}^T |\hat{\delta}_t(v)|^2$. It follows from the Cauchy–Schwarz inequality, and $|\hat{\delta}_t(v)| \leq |v| \cdot |\hat{\varepsilon}_t - \varepsilon_t|$ that

$$\begin{aligned} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{4j}(v)|^2 dW &\leq \sum_{j=1}^{T-1} k^2(j/p) \left(T_j^{-1} \sum_{t=j+1}^T \varepsilon_t \right)^2 \\ &\times \left[\sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 \right] \int v^2 dW(v) = O_P(p/T) \end{aligned}$$

given (A.3) and (A.7), and $E(\sum_{t=j+1}^T \varepsilon_t)^2 = \sigma^2 T_j$ by the *m.d.s.* hypothesis of $\{\varepsilon_t\}$. \parallel

Proof of Lemma A5. We first write

$$\hat{B}_{5j}(v) = T_j^{-1} \sum_{t=j+1}^T [\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})] \psi_{t-j}(v) + T_j^{-1} \sum_{t=j+1}^T [\varepsilon_t(\hat{\theta}) - \varepsilon_t] \psi_{t-j}(v) \equiv \hat{B}_{51j}(v) + \hat{B}_{52j}(v), \text{ say.} \tag{A.13}$$

Given $|\psi_t(v)| \leq 2$, (A.7), and Assumptions A5–A7, we have

$$\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{51j}(v)|^2 dW(v) \leq 4 \left[\sum_{t=1}^T |\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})| \right]^2 \sum_{j=1}^{T-1} a_T(j) \int dW(v) = O_P(p/T). \tag{A.14}$$

Also, by the second order Taylor series expansion in (A.11), we have

$$-\hat{B}_{52j}(v) = (\hat{\theta} - \theta_0)' T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0) \psi_{t-j}(v) + \frac{1}{2} (\hat{\theta} - \theta_0)' \left[T_j^{-1} \sum_{t=j+1}^T g''_t(\bar{\theta}) \psi_{t-j}(v) \right] (\hat{\theta} - \theta_0)$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_0 . Thus, we have

$$\begin{aligned} \sum_{T-1}^{j=1} k^2(j/p) T_j \int |\hat{B}_{52j}(v)|^2 dW(v) &\leq 2\|\sqrt{T}(\hat{\theta} - \theta_0)\|^2 \sum_{j=1}^{T-1} k^2(j/p) \\ &\times \int \left\| T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0) \psi_{t-j}(v) \right\|^2 dW(v) \\ &+ 2\|\sqrt{T}(\hat{\theta} - \theta_0)\|^4 \left[T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_0} \|g''_t(\theta)\| \right]^2 \\ &\times \left[\sum_{j=1}^{T-1} a_T(j) \right] \int dW(v) = O_P(1) + O_P(p/T), \end{aligned} \tag{A.15}$$

where the last term is $O_P(p/T)$ given (A.7), and the first term is $O_P(1)$, as is shown below:

Put $\eta_j(v) \equiv E[g'_t(\theta_0)\psi_{t-j}(v)] = \text{cov}[g'_t(\theta_0), \psi_{t-j}(v)]$. Then $\sup_{v \in \mathbb{R}} \sum_{j=1}^{\infty} \|\eta_j(v)\| \leq C$ by Assumption A8. Next, expressing the moments by cumulants via well-known formulas (e.g. Hannan, 1970, (5.1), p. 23, for real-valued processes), we can obtain

$$\begin{aligned} T_j E \left\| T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0)\psi_{t-j}(v) - \eta_j(v) \right\|^2 &\leq \sum_{\tau=-T_j}^{T_j} \|\text{cov}[g'_t(\theta_0), g'_{t-\tau}(\theta_0)']\| \cdot |\sigma_{\tau}(v, -v)| \\ &\quad + \sum_{\tau=-T_j}^{T_j} \|\eta_{j+|\tau|}(-v)\| \cdot \|\eta_{j-|\tau|}(v)\| \\ &\quad + \sum_{\tau=-T_j}^{T_j} \|\kappa_{j,|\tau|,j+|\tau|}(v)\| \leq C, \end{aligned} \quad (\text{A.16})$$

given Assumption A8, where $\kappa_{j,l,\tau}(v)$ is as in Assumption A8. See also (A.7) of Hong (1999, p. 1212). Consequently, from (A.7) and (A.16), $|k(\cdot)| \leq 1$, and $p/T \rightarrow 0$, we have

$$\begin{aligned} \sum_{j=1}^{T-1} k^2(j/p) E \int \left\| T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0)\psi_{t-j}(v) \right\|^2 dW(v) &\leq C \sum_{j=1}^{T-1} \int \|\eta_j(v)\|^2 dW(v) + C \sum_{j=1}^{T-1} a_T(j) \\ &= O(1) + O(p/T) = O(1). \end{aligned}$$

Hence the first term in (A.15) is $O_P(1)$. The desired result of Lemma A5 follows from (A.13) to (A.15). \parallel

Proof of Lemma A1. The proof is analogous to that of Lemma A4. \parallel

Proof of Proposition A2. Given the decomposition in (A.6), we have

$$|[\hat{\sigma}_j^{(1,0)}(0, v) - \tilde{\sigma}_j^{(1,0)}(0, v)]\tilde{\sigma}_j^{(1,0)}(0, v)^*| \leq \sum_{a=1}^6 |\hat{B}_{aj}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)|, \quad (\text{A.17})$$

where the $\hat{B}_{aj}(v)$ are defined in (A.6). By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{aj}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) \\ &\leq \left[\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{aj}(v)|^2 dW(v) \right]^{\frac{1}{2}} \left[\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\tilde{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) \right]^{\frac{1}{2}} \\ &= O_P(p^{\frac{1}{2}}/T^{\frac{1}{2}}) O_P(p^{\frac{1}{2}}) = o_P(p^{\frac{1}{2}}), \quad a = 1, 2, 3, 4, 6, \end{aligned} \quad (\text{A.18})$$

given Lemmas A1–A4 and A6, and $p/T \rightarrow 0$, where $p^{-1} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\tilde{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) = O_P(1)$ by Markov's inequality, the *m.d.s.* hypothesis of $\{\varepsilon_t\}$, and (A.7).

It remains to consider $a = 5$. By (A.13) and the triangular inequality, we have

$$\begin{aligned} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{5j}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) &\leq \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{51j}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) \\ &\quad + \sum_{j=1}^{T-1} k^2(j/p) T_j \\ &\quad \times \int |\hat{B}_{52j}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v). \end{aligned} \quad (\text{A.19})$$

For the first term in (A.19), we have

$$\begin{aligned} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{51j}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) &\leq 2 \left[\sum_{t=1}^T \sup_{\theta \in \Theta_0} |g(I_{t-1}^\dagger, \theta) - g(I_{t-1}, \theta)| \right] \\ &\quad \times \sum_{j=1}^{T-1} k^2(j/p) \int |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) \\ &= O_P(p/T^{\frac{1}{2}}), \end{aligned} \quad (\text{A.20})$$

given Assumption A5, (A.7), and the *m.d.s.* property of $\{\varepsilon_t\}$. For the second term in (A.19), we have

$$\begin{aligned} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{B}_{52j}(v)| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) &\leq \|\hat{\theta} - \theta_0\| \sum_{j=1}^{T-1} k^2(j/p) T_j \\ &\quad \times \int \|T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0) \psi_{t-j}(v)\| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) \\ &\quad + T \|\hat{\theta} - \theta_0\|^2 \left[T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_0} \|g''_t(\theta)\| \right] \\ &\quad \times \sum_{j=1}^{T-1} k^2(j/p) \int |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) \\ &= O_P(1 + p/T^{\frac{1}{2}}) + O_P(p/T^{\frac{1}{2}}) = o_P(p^{\frac{1}{2}}) \end{aligned} \tag{A.21}$$

given $p \rightarrow \infty$, $p/T \rightarrow 0$, and Assumptions A3–A8, where we have made use of the fact that $T_j E |\tilde{\sigma}_j^{(1,0)}(0, v)|^2 \leq C$ under the *m.d.s.* hypothesis of $\{\varepsilon_t\}$. Note that for the first term in (A.21), we have made use of the fact that

$$\begin{aligned} E \left[\left\| T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0) \psi_{t-j}(v) \right\| \left| \tilde{\sigma}_j^{(1,0)}(0, v) \right| \right] &\leq \left[E \left\| T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0) \psi_{t-j}(v) \right\|^2 \right]^{\frac{1}{2}} [E |\tilde{\sigma}_j^{(1,0)}(0, v)|^2]^{\frac{1}{2}} \\ &\leq C [\|\eta_j(v)\| + CT_j^{-\frac{1}{2}}] T_j^{-\frac{1}{2}} \end{aligned}$$

by (A.16), and consequently,

$$\begin{aligned} T^{-\frac{1}{2}} \sum_{j=1}^{T-1} k^2(j/p) T_j E \int \left\| T_j^{-1} \sum_{t=j+1}^T g'_t(\theta_0) \psi_{t-j}(v) \right\| |\tilde{\sigma}_j^{(1,0)}(0, v)| dW(v) \\ \leq C \sum_{j=1}^{T-1} \int \|\eta_j(v)\| dW(v) + CT^{-\frac{1}{2}} \sum_{j=1}^{T-1} k^2(j/p) = O(1 + p/T^{\frac{1}{2}}) \end{aligned}$$

given $|k(\cdot)| \leq 1$ and Assumption A8. Combining (A.19)–(A.21) then yields the result of this lemma. \parallel

Proof of Theorem A2. The proof is similar to that of Theorem A1. Let \hat{A}_{1q} and \hat{A}_{2q} be defined in the same way as \hat{A}_1 and \hat{A}_2 in (A.5), with $\{\varepsilon_{q,t}\}_{t=1}^T$ replacing $\{\varepsilon_t\}_{t=1}^T$. 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 $\delta_{q,t} \equiv e^{iv\varepsilon_t} - e^{iv\varepsilon_{q,t}}$ and $\psi_{q,t}(v) \equiv e^{iv\varepsilon_{q,t}} - \varphi_q(v)$, where $\varphi_q(v) \equiv E(e^{iv\varepsilon_{q,t}})$. Let $\tilde{\sigma}_{q,j}^{(1,0)}(0, v)$ be defined as $\tilde{\sigma}_j^{(1,0)}(0, v)$, with $\{\varepsilon_{q,t}\}_{t=1}^T$ replacing $\{\varepsilon_t\}_{t=1}^T$. Then, similar to (A.6), we have

$$\begin{aligned} \tilde{\sigma}_j^{(1,0)}(0, v) - \tilde{\sigma}_{q,j}^{(1,0)}(0, v) &= iT_j^{-1} \sum_{t=j+1}^T (\varepsilon_t - \varepsilon_{q,t}) \delta_{q,t-j}(v) - i \left[T_j^{-1} \sum_{t=j+1}^T (\varepsilon_t - \varepsilon_{q,t}) \right] \\ &\quad \times \left[T_j^{-1} \sum_{t=j+1}^T \delta_{q,t-j}(v) \right] + iT_j^{-1} \sum_{t=j+1}^T \varepsilon_{q,t} \delta_{q,t-j}(v) \\ &\quad - i \left(T_j^{-1} \sum_{t=j+1}^T \varepsilon_{q,t} \right) \left[T_j^{-1} \sum_{t=j+1}^T \delta_{q,t-j}(v) \right] \\ &\quad + iT_j^{-1} \sum_{t=j+1}^T (\varepsilon_t - \varepsilon_{q,t}) \psi_{q,t-j}(v) - i \left[T_j^{-1} \sum_{t=j+1}^T (\varepsilon_t - \varepsilon_{q,t}) \right] \\ &\quad \times \left[T_j^{-1} \sum_{t=j+1}^T \psi_{q,t-j}(v) \right] \\ &= i[\hat{B}_{1jq}(v) - \hat{B}_{2jq}(v) + \hat{B}_{3jq}(v) + \hat{B}_{4jq}(v) - \hat{B}_{5jq}(v) + \hat{B}_{6jq}(v)], \text{ say.} \end{aligned}$$

Following reasoning analogous to that of Theorem A1 and noting that $E(\varepsilon_t | I_{t-1}) = 0$ a.s. and $E(\varepsilon_{q,t} | I_{t-1}) = 0$ a.s., we obtain

$$p^{-\frac{1}{2}} \hat{A}_{1q} \leq 8p^{-\frac{1}{2}} \sum_{a=1}^6 \sum_{a=1}^6 k^2(j/p) T_j \int |\hat{B}_{aj,q}(v)|^2 dW(v) = O_P(p^{\frac{1}{2}}/q^\kappa) = o_P(1)$$

given Assumption A2, $q/p \rightarrow \infty$ and $\kappa \geq 1$. Moreover, by the Cauchy–Schwartz inequality, we can obtain

$$p^{-\frac{1}{2}} \hat{A}_{2q} = 2p^{-\frac{1}{2}} \sum_{a=1}^6 \sum_{j=1}^{T-1} k^2(j/p) T_j \text{Re} \int \hat{B}_{aj,q}(v) \tilde{\sigma}_{qj}^{(1,0)}(0, v)^* dW(v) = O_P(p^{\frac{1}{2}}/q^{\frac{\kappa}{2}}) = o_P(1).$$

This completes the proof of Theorem A2. \parallel

Proof of Theorem A3. We shall show Propositions A3 and A4 below.

Proposition A3. Let $\tilde{\sigma}_{q,j}^{(1,0)}(0, v)$ be defined as $\tilde{\sigma}_j^{(1,0)}(0, v)$, and let $\tilde{C}_{1q}(p)$ be defined as $\tilde{C}_1(p)$, with $\{\varepsilon_{q,t}\}_{t=1}^T$ replacing $\{\varepsilon_t\}_{t=1}^T$. Then under the conditions of Theorem 1,

$$p^{-\frac{1}{2}} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\tilde{\sigma}_{q,j}^{(1,0)}(0, v)|^2 dW(v) = p^{-\frac{1}{2}} \tilde{C}_{1q}(p) + p^{-\frac{1}{2}} \tilde{V}_q + o_p(1),$$

where $\tilde{V}_q = \sum_{t=2q+2}^T \varepsilon_{q,t} \sum_{j=1}^q a_T(j) \int \psi_{q,t-j}(v) [\sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v)] dW(v)$.

Proposition A4. Let $\tilde{D}_{1q}(p)$ be defined as $\tilde{D}_1(p)$ with $\{\varepsilon_{q,t}\}$ replacing $\{\varepsilon_t\}$. Then $\tilde{D}_{1q}^{-\frac{1}{2}}(p) \tilde{V}_q \xrightarrow{d} N(0, 1)$.

Proof of Proposition A3. Recall that $\tilde{\sigma}_{q,j}^{(1,0)}(0, v) = T_j^{-1} \sum_{t=j+1}^T \varepsilon_{q,t} \psi_{q,t}(v)$, where $\psi_{q,t}(v) \equiv e^{iv\varepsilon_{q,t}} - \varphi_q(v)$ and $\varphi_q(v) = E(e^{iv\varepsilon_{q,t}})$. We first decompose

$$\begin{aligned} \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\tilde{\sigma}_{q,j}^{(1,0)}(0, v)|^2 dW(v) &= \sum_{j=1}^{T-1} a_T(j) \int \left| \sum_{t=1}^T \varepsilon_{q,t} \psi_{q,t-j}(v) \right|^2 dW(v) \\ &\quad + \sum_{j=1}^{T-1} a_T(j) \int \left| \sum_{t=1}^j \varepsilon_{q,t} \psi_{q,t-j}(v) \right|^2 dW(v) \\ &\quad - 2\text{Re} \sum_{j=1}^{T-1} a_T(j) \int \left[\sum_{t=1}^T \varepsilon_{q,t} \psi_{q,t-j}(v) \right] \\ &\quad \times \left[\sum_{t=1}^j \varepsilon_{q,t} \psi_{q,t-j}(v) \right]^* dW(v) \\ &\equiv \tilde{Q}_q + \tilde{R}_{1q} - 2\text{Re}(\tilde{R}_{2q}). \end{aligned} \tag{A.22}$$

Next we write

$$\begin{aligned} \tilde{Q}_q &= \sum_{j=1}^{T-1} a_T(j) \int \sum_{t=1}^T \varepsilon_{q,t}^2 |\psi_{q,t-j}(v)|^2 dW(v) + 2\text{Re} \sum_{j=1}^{T-1} a_T(j) \\ &\quad \times \int \sum_{t=2}^T \sum_{s=1}^{t-1} \varepsilon_{q,t} \varepsilon_{q,s} \psi_{q,t-j}(v) \psi_{q,s-j}^*(v) \\ &\equiv \tilde{C}_q(p) + 2\text{Re}(\tilde{U}_q), \end{aligned} \tag{A.23}$$

where we further decompose

$$\begin{aligned} \tilde{U}_q &= \sum_{t=2q+2}^T \varepsilon_{q,t} \int \sum_{j=1}^{T-2} a_T(j) \psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \\ &\quad + \sum_{t=2}^T \varepsilon_{q,t} \int \sum_{j=1}^{T-2} a_T(j) \psi_{q,t-j}(v) \sum_{s=\max(1,t-2q)}^{t-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \\ &\equiv \tilde{U}_{1q} + \tilde{R}_{3q}, \end{aligned} \tag{A.24}$$

where in the first term \tilde{U}_{1q} , we have $t - s > 2q$ so that $\{\varepsilon_{q,t}, \psi_{q,t-j}(v)\}_{j=1}^q$ is independent of $\{\varepsilon_{q,s}, \psi_{q,s-j}(v)\}_{j=1}^q$ for q sufficiently large. In the second term \tilde{R}_{3q} , we have $0 < t - s \leq 2q$. Finally, we write

$$\begin{aligned} \tilde{U}_{1q} &= \sum_{t=2q+2}^T \varepsilon_{q,t} \sum_{j=1}^q a_T(j) \int \psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \\ &\quad + \sum_{t=2q+2}^T \varepsilon_{q,t} \sum_{j=q+1}^{T-1} a_T(j) \int \psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \\ &\equiv \tilde{V}_q + \tilde{R}_{4q}, \end{aligned} \tag{A.25}$$

where the first term \tilde{V}_q is contributed by the lag orders j from 1 to q ; and the second term \tilde{R}_{4q} is contributed by the lag orders $j > q$. It follows from (A.22) to (A.25) that

$$\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\tilde{\sigma}_{q,j}^{(1,0)}(0, v)|^2 dW(v) = \tilde{C}_q(p) + 2\text{Re}(\tilde{V}_q) + \tilde{R}_{1q} - 2\text{Re}(\tilde{R}_{2q} - \tilde{R}_{3q} - \tilde{R}_{4q}).$$

It suffices to show Lemmas A7–A11 below, which imply $p^{-\frac{1}{2}}[\tilde{C}_q(p) - \tilde{C}_{1q}(p)] = o_P(1)$ and $p^{-\frac{1}{2}}\tilde{R}_{aq} = o_P(1)$ for $a = 1, 2, 3, 4$ given $q = p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$ and $p = cT^\lambda$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$.

Lemma A7. Let $\tilde{C}_q(p)$ be defined as in (A.23). Then $\tilde{C}_q(p) - \tilde{C}_{1q}(p) = O_P(p^2/T)$.

Lemma A8. Let \tilde{R}_{1q} be defined as in (A.22). Then $\tilde{R}_{1q} = O_P(p^2/T)$.

Lemma A9. Let \tilde{R}_{2q} be defined as in (A.22). Then $\tilde{R}_{2q} = O_P(p^{\frac{3}{2}}/T^{\frac{1}{2}})$.

Lemma A10. Let \tilde{R}_{3q} be defined as in (A.24). Then $\tilde{R}_{3q} = O_P(q^{\frac{1}{2}}p/T^{\frac{1}{2}})$.

Lemma A11. Let \tilde{R}_{4q} be defined as in (A.25). Then $\tilde{R}_{4q} = O_P(p^{2b} \ln(T)/q^{2b-1})$.

Proof of Lemma A7. By Markov’s inequality and $E|\tilde{C}_q(p) - \tilde{C}_{1q}(p)| \leq Cp^2/T$ given $\sum_{j=1}^{T-1} (j/p)a_T(j) = O(p/T)$. \parallel

Proof of Lemma A8. By the *m.d.s.* property of $\{\varepsilon_{q,t}, \mathcal{F}_{t-1}\}$ where \mathcal{F}_{t-1} is the sigma-field generated by $\{\varepsilon_{t-j}\}_{j=1}^\infty$, we can obtain $E \int |\sum_{t=1}^j \varepsilon_{q,t} \psi_{q,t-j}(v)|^2 dW(v) = \sum_{t=1}^j \int E[\varepsilon_{q,t}^2 |\psi_{q,t-j}(v)|^2] dW(v) \leq Cj$. The result then follows from Markov’s inequality and $\sum_{j=1}^{T-1} (j/p)a_T(j) = O(p/T)$ given Assumption A6. \parallel

Proof of Lemma A9. The proof is similar to that of Lemma A8, with the fact that $E|\int[\sum_{t=1}^T \varepsilon_{q,t} \psi_{q,t-j}(v)] \times [\sum_{t=1}^j \varepsilon_{q,t} \psi_{q,t-j}(v)]^* dW(v)| \leq C(jT)^{\frac{1}{2}}$ given Assumption A6. \parallel

Proof of Lemma A10. By the *m.d.s.* property of $\{\varepsilon_{q,t}, \mathcal{F}_{t-1}\}$, Minkowski’s inequality and (A.7), we have

$$\begin{aligned} E|\tilde{R}_{3q}|^2 &= \sum_{t=2}^T E \left| \sum_{j=1}^{T-1} a_T(j) \int \varepsilon_{q,t} \psi_{q,t-j}(v) \sum_{s=\max(1,t-2q)}^{t-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \right|^2 \\ &\leq \sum_{t=2}^T \left[\sum_{j=1}^{T-1} a_T(j) \int \left(E \left| \varepsilon_{q,t} \psi_{q,t-j}(v) \sum_{s=\max(1,t-2q)}^{t-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) \right|^2 \right)^{\frac{1}{2}} dW(v) \right]^2 \\ &\leq 2CTq \left[\sum_{j=1}^{T-1} a_T(j) \right]^2 = O(qp^2/T). \end{aligned}$$

Proof of Lemma A11. By the *m.d.s.* property of $\{\varepsilon_{q,t}, \mathcal{F}_{t-1}\}$ and Minkowski’s inequality, we have

$$\begin{aligned} E|\tilde{R}_{4q}|^2 &= \sum_{t=2q+2}^T E \left| \sum_{j=q+1}^{T-1} a_T(j) \int \varepsilon_{q,t} \psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \right|^2 \\ &\leq \sum_{t=2q+2}^T \left[\sum_{j=q+1}^{T-1} a_T(j) \int \left(E \left| \varepsilon_{q,t} \psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) \right|^2 \right)^{\frac{1}{2}} dW(v) \right]^2 \\ &\leq CT^2 \left[\sum_{j=q+1}^{T-1} a_T(j) \right]^2 \leq C^3 T^2 \left[\sum_{j=q+1}^{T-1} (j/p)^{-2b} T_j^{-1} \right]^2 = O(p^{4b} \ln^2(T)/q^{4b-2}) \end{aligned}$$

given Assumption A6 (*i.e.* $k(z) \leq C|z|^{-b}$ as $z \rightarrow \infty$). \parallel

Proof of Proposition A4. We rewrite $\tilde{V}_q = \sum_{t=2q+2}^T V_q(t)$, where

$$V_q(t) = \varepsilon_{q,t} \sum_{j=1}^q a_T(j) \int \psi_{q,t-j}(v) H_{j,t-2q-1}(v) dW(v),$$

and $H_{j,t-2q-1}(v) = \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v)$. We apply the martingale limit theorem (Brown, 1971), which states $\text{var}(2\text{Re } \tilde{V}_q)^{-\frac{1}{2}} 2\text{Re } \tilde{V}_q \xrightarrow{d} N(0, 1)$ if

$$\text{var}(2\text{Re } \tilde{V}_q)^{-1} \sum_{t=1}^T [2\text{Re } V_q(t)]^2 \mathbf{1}[|2\text{Re } V_q(t)| > \eta \cdot \text{var}(2\text{Re } \tilde{V}_q)^{\frac{1}{2}}] \rightarrow 0 \quad \forall \eta > 0, \tag{A.26}$$

$$\text{var}(2\text{Re } \tilde{V}_q)^{-1} \sum_{t=1}^T E \left[2\text{Re } V_q^2(t) | \mathcal{F}_{t-1} \right] \xrightarrow{p} 1. \tag{A.27}$$

First, we compute $\text{var}(2\text{Re } \tilde{V}_q)$. By the *m.d.s.* property of $\{\varepsilon_{q,t}, \mathcal{F}_{t-1}\}$ under \mathbb{H}_0 and independence between $\varepsilon_{q,t}$ and $\{\varepsilon_{t-j-1}\}_{j=q}^\infty$ for q sufficiently large, we have

$$\begin{aligned} E(\tilde{V}_q^2) &= \sum_{t=2q+2}^T E \left[\varepsilon_{q,t}^2 \left(\int \sum_{j=1}^q a_T(j) \psi_{q,t-j}(v) \sum_{s=1}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}^*(v) dW(v) \right)^2 \right] \\ &= \sum_{j=1}^q \sum_{l=1}^q a_T(j) a_T(l) \iint \sum_{t=2q+2}^T \sum_{s=1}^{t-2q-1} E \left[\varepsilon_{q,t}^2 \psi_{q,t-j}(v) \psi_{q,t-l}(v') \right] \\ &\quad \times E \left[\varepsilon_{q,s}^2 \psi_{q,s-j}^*(v) \psi_{q,s-l}^*(v') \right] dW(v) dW(v') \\ &= \frac{1}{2} \sum_{j=1}^q \sum_{l=1}^q k^2(j/p) k^2(l/p) \iint |E[\varepsilon_{q,0}^2 \psi_{q,-j}(v) \psi_{q,-l}(v')]|^2 dW(v) dW(v') [1 + o(1)]. \end{aligned}$$

Similarly, we can obtain

$$\begin{aligned} E(\tilde{V}_q^*)^2 &= \frac{1}{2} \sum_{j=1}^q \sum_{l=1}^q k^2(j/p) k^2(l/p) \iint |E[\varepsilon_{q,0}^2 \psi_{q,-j}(v) \psi_{q,-l}(v')]|^2 dW(v) dW(v') [1 + o(1)], \\ E|\tilde{V}_q|^2 &= \frac{1}{2} \sum_{j=1}^q \sum_{l=1}^q k^2(j/p) k^2(l/p) \iint |E[\varepsilon_{q,0}^2 \psi_{q,-j}(v) \psi_{q,-l}^*(v')]|^2 dW(v) dW(v') [1 + o(1)]. \end{aligned}$$

Because $W(\cdot)$ weighs sets symmetric about zero equally, we have $E|\tilde{V}_q|^2 = E(\tilde{V}_q^2) = E(\tilde{V}_q^*)^2$. Hence,

$$\begin{aligned} \text{var}(2\text{Re } \tilde{V}_q) &= E(\tilde{V}_q^2) + E(\tilde{V}_q^*)^2 + 2E|\tilde{V}_q|^2 \\ &= 2 \sum_{j=1}^q \sum_{l=1}^q k^2(j/p) k^2(l/p) \iint |E[\varepsilon_0^2 \psi_{-j}(v) \psi_{-l}(v')]|^2 dW(v) dW(v') [1 + o(1)], \tag{A.28} \end{aligned}$$

where we have made use of the fact that $E[\varepsilon_{q,0}^2 \psi_{q,-j}(v) \psi_{q,-l}^*(v')] \rightarrow E[\varepsilon_0^2 \psi_{-j}(v) \psi_{-l}(v')]$ as $q \rightarrow \infty$ given Assumption A2. Put $C(0, j, l) \equiv E[(\varepsilon_0^2 - \sigma^2) \psi_{-j}(v) \psi_{-l}(v')]$. Then

$$\begin{aligned} E[\varepsilon_0^2 \psi_{-j}(v) \psi_{-l}(v')] &= C(0, j, l) + \sigma^2 \sigma_{l-j}(v, v'), \\ |E[\varepsilon_0^2 \psi_{-j}(v) \psi_{-l}(v')]|^2 &= |C(0, j, l)|^2 + \sigma^4 |\sigma_{l-j}(v, v')|^2 + 2\sigma^2 \text{Re}[C(0, j, l) \sigma_{l-j}^*(v, v')]. \end{aligned}$$

Given $\sum_{j=-\infty}^\infty \sum_{l=-\infty}^\infty |C(0, j, l)| \leq C$, and $|k(\cdot)| \leq 1$, we have

$$\begin{aligned} \text{var}(2\text{Re } \tilde{V}_q) &= 2\sigma^4 \sum_{j=1}^q \sum_{l=1}^q k^2(j/p) k^2(l/p) \iint |\sigma_{l-j}(v, v')|^2 dW(v) dW(v') [1 + o(1)] \\ &= 2\sigma^4 p \sum_{m=1-q}^{q-1} \left[p^{-1} \sum_{j=m+1}^q k^2(j/p) k^2[(j-m)/p] \right] \\ &\quad \times \iint |\sigma_m(v, v')|^2 dW(v) dW(v') [1 + o(1)] \\ &= 2\sigma^4 p \int_0^\infty k^4(z) dz \sum_{m=-\infty}^\infty \iint |\sigma_m(v, v')|^2 dW(v) dW(v') [1 + o(1)] \\ &= 4\pi\sigma^4 p \int_0^\infty k^4(z) dz \int \int_{-\pi}^\pi |f(\omega, v, v')|^2 d\omega dW(v) dW(v') [1 + o(1)], \end{aligned}$$

where we used the fact that for any given m , $p^{-1} \sum_{j=m+1}^q k^2(j/p) k^2(\frac{j-m}{p}) \rightarrow \int_0^\infty k^4(z) dz$ as $p \rightarrow \infty, q/p \rightarrow 0$.

We now verify condition (A.26). Noting that $E|H_{j,t-2q-1}(v)|^4 \leq Ct^2$ for $1 \leq j \leq q$ given the *m.d.s.* property of $\{\varepsilon_{q,t}, \mathcal{F}_{t-1}\}$ and Rosenthal's inequality (cf. Hall and Heyde, 1980, p. 23), we have

$$\begin{aligned} E|V_q(t)|^4 &\leq \left[\sum_{j=1}^q a_T(j) \int \left(E|\varepsilon_{q,t}\psi_{q,t-j}(v)H_{j,t-2q-1}(v)|^4 \right)^{\frac{1}{4}} dW(v) \right]^4 \\ &\leq Ct^2 \left[\sum_{j=1}^q a_T(j) \right]^4 = O(p^4t^2/T^4). \end{aligned}$$

It follows that $\sum_{t=2q+2}^T E|V_q(t)|^4 = O(p^4/T) = o(p^2)$ given $p^2/T \rightarrow 0$. Thus, (A.26) holds.

Next, we verify condition (A.27). Put $\sigma_{q,t}^2 \equiv E(\varepsilon_{q,t}^2 | \mathcal{F}_{t-1})$. Then

$$\begin{aligned} E[V_q^2(t)|\mathcal{F}_{t-1}] &= \sigma_{q,t}^2 \left[\sum_{j=1}^q a_T(j) \int \psi_{q,t-j}(v)H_{j,t-2q-1}(v) \right]^2 \\ &= \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \iint \sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v') \\ &\quad \times H_{j,t-2q-1}(v)H_{l,t-2q-1}(v')dW(v)dW(v') \\ &= \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \iint E[\sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')] \\ &\quad \times H_{j,t-2q-1}(v)H_{l,t-2q-1}(v')dW(v)dW(v') \\ &\quad + \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \iint \tilde{Z}_{q,t}^{j,l}(v, v')H_{j,t-2q-1}(v)H_{l,t-2q-1}(v')dW(v)dW(v') \\ &\equiv S_{1q}(t) + V_{1q}(t), \text{ say.} \end{aligned} \tag{A.29}$$

where $\tilde{Z}_{q,t}^{j,l}(v, v') \equiv \sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v') - E[\sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')]$. We further decompose

$$\begin{aligned} S_{1q}(t) &= \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E[\sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v)]E[H_{j,t-2q-1}(v)H_{l,t-2q-1}(v')]dW(v)dW(v') \\ &\quad + \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E[\sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')] \\ &\quad \times \{H_{j,t-2q-1}(v)H_{l,t-2q-1}(v') - E[H_{j,t-2q-1}(v)H_{l,t-2q-1}(v')]\}dW(v)dW(v') \\ &\equiv E[V_q^2(t)] + S_{2q}(t), \text{ say,} \end{aligned} \tag{A.30}$$

where

$$E[V_q^2(t)] = \sum_{j=1}^q \sum_{l=1}^q (t-q-1)a_T(j)a_T(l) \int |E[\sigma_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')]|dW(v)dW(v').$$

Put $Z_{q,s}^{j,l}(v, v') \equiv \varepsilon_{q,s}^2 \psi_{q,s-j}(v)\psi_{q,s-l}(v') - E[\varepsilon_{q,s}^2 \psi_{q,s-j}(v)\psi_{q,s-l}(v')]$. Then we write

$$\begin{aligned} S_{2q}(t) &= \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E[\varepsilon_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')] \sum_{s=1}^{t-2q-1} Z_{q,s}^{j,l}(v, v')dW(v)dW(v') \\ &\quad + \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E[\varepsilon_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')] \\ &\quad \times \sum_{s=2}^{t-2q-1} \sum_{\tau=1}^{s-1} \varepsilon_{q,s} \psi_{q,s-j}(v)\varepsilon_{q,\tau} \psi_{q,\tau-l}(v')dW(v)dW(v') \\ &\equiv V_{2q}(t) + S_{3q}(t), \text{ say,} \end{aligned} \tag{A.31}$$

where

$$\begin{aligned} S_{3q}(t) &= \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E[\varepsilon_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')] \\ &\quad \times \sum_{0 < s-\tau \leq 2q} \sum_{0 < s-\tau \leq 2q} \varepsilon_{q,s} \psi_{q,s-j}(v)\varepsilon_{q,\tau} \psi_{q,\tau-l}(v')dW(v)dW(v') \\ &\quad + \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E[\varepsilon_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')] \\ &\quad \times \sum_{s-\tau > 2q} \sum_{s-\tau > 2q} \varepsilon_{q,s} \psi_{q,s-j}(v)\varepsilon_{q,\tau} \psi_{q,\tau-l}(v')dW(v)dW(v') \\ &\equiv V_{3q}(t) + V_{4q}(t), \text{ say.} \end{aligned} \tag{A.32}$$

It follows from (A.29) to (A.32) that $\sum_{t=2q+2}^T \{E[V_q^2(t)|\mathcal{F}_{t-1}] - E[V_q^2(t)]\} = \sum_{a=1}^4 \sum_{t=2q+2}^T V_{aq}(t)$. It suffices to show Lemmas A12–A15 below, which imply $E|\sum_{t=2q+2}^T E[V_q^2(t)|\mathcal{F}_{t-1}] - E[V_q^2(t)]|^2 = o(p^2)$ given $q = p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$ and $p = cT^\lambda$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$. Thus, condition (A.27) holds, and so $M_{1q}(p) \xrightarrow{d} N(0, 1)$ by Brown’s theorem.

Lemma A12. *Let $V_{1q}(t)$ be defined as in (A.29). Then $E|\sum_{t=2q+2}^T V_{1q}(t)|^2 = O(qp^4/T)$.*

Lemma A13. *Let $V_{2q}(t)$ be defined as in (A.31). Then $E|\sum_{t=2q+2}^T V_{2q}(t)|^2 = O(qp^4/T)$.*

Lemma A14. *Let $V_{3q}(t)$ be defined as in (A.32). Then $E|\sum_{t=2q+2}^T V_{3q}(t)|^2 = O(qp^4/T)$.*

Lemma A15. *Let $V_{4q}(t)$ be defined as in (A.32). Then $E|\sum_{t=2q+2}^T V_{4q}(t)|^2 = O(p)$.*

Proof of Lemma A12. Recall the definition of $\tilde{Z}_{q,t}^{j,l}(v, v')$ as in (A.29). Noting that $\tilde{Z}_{q,t}^{j,l}(v, v')$ is independent of $\{H_{j,t-2q-1}(v)H_{l,t-2q-1}(v')\}$ and that $\tilde{Z}_{q,t}^{j,l}(v, v')$ is independent of $\tilde{Z}_{q,\tau}^{j,l}(v, v')$ for $t - \tau > 2q$ and $1 \leq j, l \leq q$, we can obtain

$$E \left| \sum_{t=2q+2}^T \tilde{Z}_{q,t}^{j,l}(v, v') H_{j,t-2q-1}(v) H_{l,t-2q-1}(v') \right|^2 \leq \sum_{|t-\tau| \leq 2q} \sum_{|t-\tau| \leq 2q} E|\tilde{Z}_{q,t}^{j,l}(v, v') \tilde{Z}_{q,\tau}^{j,l}(v, v')| \times (E|H_{j,t-2q-1}(v)|^4)^{\frac{1}{4}} (E|H_{l,t-2q-1}(v')|^4)^{\frac{1}{4}} \times (E|H_{j,\tau-2q-1}(v)|^4)^{\frac{1}{4}} (E|H_{l,\tau-2q-1}(v')|^4)^{\frac{1}{4}} = O(T^3 q),$$

where we have made use of the fact that $E|H_{j,t-2q-1}(v)|^4 \leq Ct^2$ for $1 \leq j \leq q$. It follows by Minkowski’s inequality and (A.7) that

$$E \left| \sum_{t=2q+2}^T V_{1q}(t) \right|^2 \leq \left[\sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \left(E \left| \sum_{t=2q+2}^T \iint \tilde{Z}_{q,t}^{j,l}(v, v') H_{j,t-2q-1}(v) H_{l,t-2q-1}(v') dW dW' \right|^2 \right)^{\frac{1}{2}} \right]^2 = O(qp^4/T).$$

Proof of Lemma A13. Recalling the definition of $Z_{q,s}^{j,l}(v, v')$ in (A.31) and noting that $\{Z_{q,s}^{j,l}(v, v')\}_{j,l=1}^q$ is independent of $\{Z_{q,\tau}^{j,l}(v, v')\}_{j,l=1}^q$ for $|s - \tau| > 2q$ where q is sufficiently large, we have $E|\sum_{s=1}^{t-2q-1} Z_{q,s}^{j,l}(v, v')|^2 = \Sigma \sum_{|s-\tau| \leq 2q} E[Z_{q,s}^{j,l}(v, v') Z_{q,\tau}^{j,l}(v, v')] \leq 2Ctq$. It follows that

$$E \left| \sum_{t=2q+2}^T V_{2q}(t) \right|^2 \leq \left\{ \sum_{t=2q+2}^T [E|V_{2q}(t)|^2]^{\frac{1}{2}} \right\}^2 \leq \left\{ \sum_{t=2q+2}^T \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int |E[e_{q,t}^2 \psi_{q,t-j}(v) \psi_{q,t-l}(v')]| \times \left(E \left| \sum_{s=1}^{t-2q-1} Z_{q,s}^{j,l}(v, v') \right|^2 \right)^{\frac{1}{2}} dW(v)dW(v') \right\}^2 = O(qp^4/T). \quad \parallel$$

Proof of Lemma A14. The result that $E|\sum_{t=2q+2}^T V_{3q}(t)|^2 = O(qp^2/T)$ by Minkowski's inequality and

$$\begin{aligned}
 E|V_{3q}(t)|^2 &\leq \left[\sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int |E[\varepsilon_{q,t}^2 \psi_{q,t-j}(v)\psi_{q,t-l}(v')]| \right. \\
 &\quad \left. \times \left[\sum_{s=1}^{t-2q-1} E \left| \varepsilon_{q,s} \psi_{q,s-j}(v) \sum_{s-\tau \leq 2q} \varepsilon_{q,\tau} \psi_{q,\tau-l}(v') \right|^2 \right]^{\frac{1}{2}} dW(v)dW(v') \right]^2 \\
 &\leq 2Ctq \left[\sum_{j=1}^q a_T(j) \right]^4 = O(tqp^4/T^4). \quad \parallel
 \end{aligned}$$

Proof of Lemma A15. The result that $E|\sum_{t=2q+2}^T V_{4q}(t)|^2 = O(p)$ follows from Minkowski's inequality, $p \rightarrow \infty$, and the fact that

$$\begin{aligned}
 E|V_{4q}(t)|^2 &= E \left| \sum_{j=1}^q \sum_{l=1}^q a_T(j)a_T(l) \int E \left[\varepsilon_{q,0}^2 \psi_{q,-j}(v)\psi_{q,-l}(v') \right] \sum_{s=2q+2}^{t-2q-1} \varepsilon_{q,s} \psi_{q,s-j}(v) \right. \\
 &\quad \left. \times \sum_{\tau=1}^{s-2q-1} \varepsilon_{q,\tau} \psi_{q,\tau-l}(v') dW dW' \right|^2 \\
 &= \sum_{j_1=1}^q \sum_{j_2=1}^q \sum_{l_1=1}^q \sum_{l_2=1}^q a_T(j_1)a_T(j_2)a_T(l_1)a_T(l_2) \iiint E[\varepsilon_{q,0}^2 \psi_{q,-j_1}(v_1)\psi_{q,-l_1}(v'_1)] \\
 &\quad \times E[\varepsilon_{q,0}^2 \psi_{q,-j_2}^*(v_2)\psi_{q,-l_2}^*(v'_2)] \sum_{s=2q+2}^{t-2q-1} E[\varepsilon_{q,s}^2 \psi_{q,s-j_1}(v_1)\psi_{q,s-j_2}(v_2)] \\
 &\quad \times \sum_{\tau=1}^{s-2q-1} E[\varepsilon_{q,\tau}^2 \psi_{q,\tau-l_1}(v'_1)\psi_{q,\tau-l_2}^*(v'_2)] \\
 &\quad \times dW(v_1)dW(v'_1)dW(v_2)dW(v'_2) = O(t^2 p/T^4)
 \end{aligned}$$

given Assumptions A2 and A8 (i.e. $\sum_{j=1}^\infty |\sigma_j(u, v)| \leq C$ and $\sum_{j=1}^\infty \sum_{l=1}^\infty |E[(\varepsilon_0^2 - \sigma^2)\psi_{-j}(u)\psi_{-l}(v)]| \leq C$). \parallel

Proof of Theorem 2. We consider $\hat{M}_1(p)$ only. The proof of Theorem 2 consists of the proofs of Theorems A4 and A5 below.

Theorem A4. Under the conditions of Theorem 2, $(p^{\frac{1}{2}}/T)[\hat{M}_1(p) - M_1(p)] \xrightarrow{P} 0$.

Theorem A5. Under the conditions of Theorem 2,

$$(p^{\frac{1}{2}}/T)M_1(p) \xrightarrow{P} (2D)^{-\frac{1}{2}} \pi \int_{-\pi}^{\pi} |f^{(0,1,0)}(\omega, 0, v) - f_0^{(0,1,0)}(\omega, 0, v)|^2 d\omega dW(v).$$

Proof of Theorem A4. It suffices to show that

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/p)T_j \left[|\hat{\sigma}_j^{(1,0)}(0, v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0, v)|^2 \right] dW(v) \xrightarrow{P} 0, \tag{A.33}$$

$p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)] = O_P(1)$, and $p^{-1}[\hat{D}_1(p) - \tilde{D}_1(p)] \xrightarrow{P} 0$, where $\tilde{C}_1(p)$ and $\tilde{D}_1(p)$ are defined in the same way as $\hat{C}_1(p)$ and $\hat{D}_1(p)$ in (3.11), with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. Since the proofs for $p^{-1}[\hat{C}_1(p) - \tilde{C}_1(p)] = O_P(1)$ and $p^{-1}[\hat{D}_1(p) - \tilde{D}_1(p)] \xrightarrow{P} 0$ are straightforward, we focus on the proof of (A.33). From (A.7), the Cauchy-Schwarz inequality, and the fact that $T^{-1} \int \sum_{j=1}^{T-1} k^2(j/p)T_j |\tilde{\sigma}_j^{(1,0)}(0, v)|^2 dW(v) = O_P(1)$ as is implied by Theorem A5 (the proof of Theorem A5 does not depend on Theorem A4, it suffices to show that $T^{-1} \hat{A}_1 \xrightarrow{P} 0$, where \hat{A}_1 is defined as in (A5). Given (A6), we shall show that $T^{-1} \int \sum_{j=1}^{T-1} k^2(j/p)T_j |\hat{B}_{aj}(v)|^2 dW(v) \xrightarrow{P} 0$, $a = 1, \dots, 6$. We first consider $a = 1$. By the Cauchy-Schwarz inequality and $|\hat{\delta}_t(v)| \leq 2$, we have

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

It follows from (A.3) and (A.7) and Assumption A7 that

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

The proof for $a = 2$ is similar, noting that $|T_j^{-1} \sum_{t=j+1}^T (\hat{\varepsilon}_t - \varepsilon_t)|^2 \leq T_j^{-1} \sum_{t=j+1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2$.

Next, we consider $a = 3$. By the Cauchy–Schwarz inequality, we have

$$|B_{3j}(v)|^2 \leq \left(T_j^{-1} \sum_{t=1}^T \varepsilon_t^2 \right) T_j^{-1} \sum_{t=j+1}^T |\hat{\delta}_{t-j}(v)|^2 \leq v^2 \left(T_j^{-1} \sum_{t=1}^T \varepsilon_t^2 \right) T_j^{-1} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2.$$

It follows that

$$\begin{aligned} T^{-1} \int \sum_{j=1}^{T-1} k^2(j/p) T_j |\hat{B}_{3j}(v)|^2 dW(v) &\leq \left(T^{-1} \sum_{t=1}^T \varepsilon_t^2 \right) \left[T^{-1} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 \right] \\ &\quad \times \sum_{j=1}^{T-1} k^2(j/p) \int v^2 dW(v) \\ &= O_P(p/T). \end{aligned}$$

The proof for $a = 4, 5, 6$ is similar to that for $a = 3$, noting that $|T_j^{-1} \sum_{t=j+1}^T \hat{\delta}_t(v)|^2 \leq T_j^{-1} \sum_{t=j+1}^T |\hat{\delta}_t(v)|^2$. This completes the proof for Theorem A4. \parallel

Proof of Theorem A5. The proof is very similar to Hong (1999, Proof of Thm. 5), for the case $(m, l) = (1, 0)$ and $W_1(\cdot) = \delta(\cdot)$, the Dirac delta function. \parallel

Proof of Theorem 3. Again, we only consider $\hat{M}_1(p)$. We shall show Theorems A6 and A7 below.

Theorem A6. Under the conditions of Theorem 3, $\hat{M}_1(\hat{p}) - M_1(\hat{p}) \xrightarrow{P} 0$.

Theorem A7. Under the conditions of Theorem 3, $M_1(\hat{p}) - M_1(p) \xrightarrow{P} 0$.

Proof of Theorem A6. Put $\hat{B} \equiv \sum_{j=1}^{T-1} k^2(j/\hat{p}) T_j \int [|\hat{\sigma}_j^{(1,0)}(0, v)|^2 - |\tilde{\sigma}_j^{(1,0)}(0, v)|^2] dW(v)$. 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 shall show $p^{-\frac{1}{2}} \hat{B} \xrightarrow{P} 0$; the proof of $p^{-\frac{1}{2}} [\hat{C}_1(\hat{p}) - \tilde{C}_1(\hat{p})] \xrightarrow{P} 0$ and $p^{-1} [\hat{D}(\hat{p}) - \tilde{D}(\hat{p})] \xrightarrow{P} 0$ is 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 A6. It follows that for any constants $\epsilon, \eta > 0$,

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

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 $T \rightarrow \infty$.

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

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

for any $\eta > 0$ given (A.6), where the inequality follows from the fact that $|k(z)| \leq k_0(z)$. This completes the proof of Theorem A6. \parallel

Proof of Theorem A7. Put $Q(p) \equiv 2\pi \sum_{j=1}^{T-1} k^2(j/\hat{p}) T_j \int |\tilde{\sigma}_j^{(1,0)}(0, v)|^2 dW(v)$. Then we can write

$$\begin{aligned} M_1(\hat{p}) - M_1(p) &= \left[\tilde{Q}(\hat{p}) - \tilde{C}_1(\hat{p}) \right] / \sqrt{\tilde{D}(\hat{p})} - \left[\tilde{Q}(p) - \tilde{C}_1(p) \right] / \sqrt{\tilde{D}(p)} \\ &= \{ [\tilde{Q}(\hat{p}) - \tilde{Q}(p)] - [\tilde{C}_1(\hat{p}) - \tilde{C}_1(p)] \} / \sqrt{\tilde{D}(\hat{p})} + M_1(p) \left[\sqrt{\tilde{D}(\hat{p})/\tilde{D}(p)} - 1 \right]. \end{aligned}$$

Following a reasoning analogous to the proof of Hong (1999, Thm. 4), we can obtain $p^{-\frac{1}{2}} [\tilde{Q}(\hat{p}) - \tilde{Q}(p)] \xrightarrow{P} 0$ under \mathbb{H}_0 . This and Lemma A16 below imply $M_1(\hat{p}) - M_1(p) \xrightarrow{P} 0$. Hence, $M_1(\hat{p}) \xrightarrow{d} N(0, 1)$. \parallel

Lemma A16. Suppose Assumptions A6 and A8 hold. If $\hat{p}/p = 1 + O_p(p^{-(\frac{3\beta}{2}-1)})$ for some $\beta > 1 + \frac{1}{4b-2}$, where b is as in Assumption A6 and $p = cn^\lambda$ for $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$ and $0 < c < \infty$. Then $p^{-\frac{1}{2}}[\hat{C}_1(\hat{p}) - \tilde{C}_1(p)] \xrightarrow{P} 0$ and $p^{-1}[\hat{D}(\hat{p}) - \tilde{D}(p)] \xrightarrow{P} 0$.

Proof of Lemma A16. The proof is analogous to the proof of Lemma A.2 in Hong (1999, pp. 1217–1218). Note that the factors $\int T_j \sum_{t=j+1}^T \varepsilon_t^2 |\psi_{t-j}(v)|^2 dW(v)$ and $\int T_j \sum_{t=\max(j,l)+1}^T \varepsilon_t^2 |\psi_{t-j}(v)\psi_{t-l}(v')|^2 dW(v)dW(v')$ in $\tilde{C}_1(p)$ and $\tilde{D}_1(p)$, respectively, do not appear in Lemma A.2 of Hong (1999), but this does not alter the proof much, provided that we apply Markov's inequality to obtain the orders of magnitude for the terms involving these stochastic factors. Note also that the condition $\lambda \in (0, 1)$ as imposed in Hong (1999) is implied by the condition that $0 < \lambda < (3 + \frac{1}{4b-2})^{-1}$. \parallel

Acknowledgements. We thank two referees and the editors (Hidehiko Ichimura and Bernard Salanié) for careful and insightful comments. We also thank A. R. Gallant, Jiti Gao, Eric Ghysels, Alastair Hall, Jerry Hausman, Han Hong, Chang-Jin Kim, Joon Park, George Tauchen, Barbara Rossi, Yoon-Jae Whang, and seminar participants at the 2003 North American Econometrics Society summer meeting and the Triangle Econometrics workshop for helpful comments, discussions and references, and the National Science Foundation for support via NSF Grant SES-0111769.

REFERENCES

- BARNETT, W., GALLANT, R., HINICH, M., JUNGEILGES, J., KAPLAN, D. and JENSEN, M. (1997), "A Single-Blind Controlled Competition Among Tests for Nonlinearity and Chaos", *Journal of Econometrics*, **82**, 157–192.
- BERA, A. K. and HIGGINS, M. L. (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence", *Journal of Business and Economic Statistics*, **15**, 43–50.
- BIERENS, H. (1982), "Consistent Model Specification Tests", *Journal of Econometrics*, **20**, 105–134.
- BOUDOUGH, J., RICHARDSON, M. and WHITELAW, R. (1994), "A Tale of Three Schools: Insights on Autocorrelations of Short Horizon Returns", *Review of Financial Studies*, **7**, 539–573.
- BRILLINGER, D. and ROSENBLATT, M. (1967a), "Asymptotic Theory of Estimates of k -th Order Spectra", in B. Harris (ed.) *Spectral Analysis of Time Series* (New York: John Wiley) 153–188.
- BRILLINGER, D. and ROSENBLATT, M. (1967b), "Computation and Interpretation of k -th Order Spectra", in B. Harris (ed.) *Spectral Analysis of Time Series* (New York: John Wiley) 189–232.
- BROCK, W., DECHERT, D., SCHEINKMAN, J. and LEBARON, B. (1996), "A Test for Independence Based on the Correlation Dimension", *Econometric Reviews*, **15**, 197–235.
- BROCK, W., HSIEH, D. and LEBARON, B. (1991) *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence* (Cambridge, MA: MIT Press).
- BROWN, B. M. (1971), "Martingale Central Limit Theorems", *Annals of Mathematical Statistics*, **42**, 59–66.
- CAMPBELL, C., LO, A. and MACKINLAY, A. C. (1997) *Econometrics of Financial Markets* (Princeton, NJ: Princeton University Press).
- CHACKO, G. and VICEIRA, L. M. (2003), "Spectral GMM Estimation of Continuous-Time Processes", *Journal of Econometrics*, **116**, 259–292.
- CHEN, S. X., HÄRDLE, W. and LI, M. (2003), "An Empirical Likelihood Goodness-of-Fit Test for Time Series", *Journal of Royal Statistical Society, Series B*, **65**, 663–678.
- COCHRANE, J. (1988), "How Big is the Random Walk in GNP?", *Journal of Political Economy*, **96**, 893–920.
- COCHRANE, J. (2001) *Asset Pricing* (Princeton, NJ: Princeton University Press).
- DAHL, C. (2002), "An Investigation of Tests for Linearity and the Accuracy of Likelihood Based Inference Using Random Fields", *Econometric Journal*, **5**, 263–284.
- DAHL, C. and GONZALEZ-RIVERA, G. (2003), "Testing for Neglected Nonlinearity in Regression Models Based on the Theory of Random Fields", *Journal of Econometrics*, **114**, 141–164.
- DIEBOLD, F. X. (1986), "Testing for Serial Correlation in the Presence of ARCH", in *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 323–328.
- DIEBOLD, F. X. and NASON, J. (1990), "Nonparametric Exchange Rate Prediction", *Journal of International Economics*, **28**, 315–332.
- EITRHEIM, Ø. and TERÄSVIRTA, T. (1996), "Testing the Adequacy of Smooth Transition Autoregressive Models", *Journal of Econometrics*, **74**, 59–76.
- ENGLE, R., LILIE, D. and ROBINS, R. (1987), "Estimating Time Varying Risk Premium in the Term Structure: The ARCH-M Model", *Econometrica*, **55**, 391–407.
- EPPS, T. (1987), "Testing that a Stationary Time Series is Gaussian", *Annals of Statistics*, **15**, 1683–1698.
- EPPS, T. (1988), "Testing that a Gaussian Process is Stationary", *Annals of Statistics*, **16**, 1667–1683.
- FAMA, E. F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, **25**, 383–417.
- FAMA, E. F. (1991), "Efficient Markets II", *Journal of Finance*, **46**, 1575–1618.
- GALLANT, A. R., HSIEH, D. and TAUCHEN, G. (1991), "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974–83", in W. Barnett, J. Powell and G. Tauchen (eds.) *Nonparametric and Semiparametric Methods*

- in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics* (Cambridge: Cambridge University Press) 199–240.
- GAO, J. and KING, M. (2001), “Estimation and Model Specification Testing in Nonparametric and Semiparametric Time Series Econometric Models” (Working Paper, School of Mathematics and Statistics, University of Western Australia).
- GAO, J., TONG, H. and WOLFF, R. (2002), “Model Specification Tests in Nonparametric Stochastic Regression Models”, *Journal of Multivariate Analysis*, **83**, 324–359.
- GLOSTEN, L., JAGANNATHAN, R. and RUNKLE, D. (1993), “On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”, *Journal of Finance*, **48**, 1779–1801.
- GRANGER, C. W. J. (1992), “Forecasting Stock Market Prices: Lessons for Forecasters”, *International Journal of Forecasting*, **8**, 3–13.
- GRANGER, C. W. J. (1995), “Modeling Nonlinear Relationships Between Extended Memory Variables”, *Econometrica*, **63**, 265–593.
- GRANGER, C. W. J. (2001), “Overview of Nonlinear Macroeconometric Empirical Models”, *Macroeconomic Dynamics*, **5**, 466–481.
- GRANGER, C. W. J. and TERÄSVIRTA, T. (1993) *Modelling Nonlinear Economic Relationships* (New York: Oxford University Press).
- HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and Its Application* (New York: Academic Press).
- HALL, R. E. (1978), “Stochastic Implications of the Life Cycle—Permanent Income Hypothesis: Theory and Evidence”, *Journal of Political Economy*, **86**, 971–987.
- HAMILTON, J. D. (1989), “A New Approach to Economic Analysis of Nonstationary Time Series and the Business Cycle”, *Econometrica*, **57**, 357–384.
- HAMILTON, J. D. (2001), “A Parametric Approach to Flexible Nonlinear Inference”, *Econometrica*, **69**, 537–573.
- HAMILTON, J. D. and LIN, G. (1996), “Stock Market Volatility and the Business Cycles”, *Journal of Applied Econometrics*, **11**, 573–593.
- HANNAN, E. (1970) *Multiple Time Series* (New York: Wiley).
- HANNAN, E. (1980), “The Estimation of the Order of an ARMA Process”, *Annals of Statistics*, **8**, 1071–1081.
- HANSEN, B. (1994), “Autoregressive Conditional Density Estimation”, *International Economic Review*, **35**, 705–730.
- HANSEN, B. (1996), “Inference when a Nuisance Parameter is not Identified under the Null Hypothesis”, *Econometrica*, **64**, 413–430.
- HANSEN, B. (1999), “Testing for Linearity”, *Journal of Economic Surveys*, **13**, 551–576.
- HANSEN, B. (2000), “Sample Splitting and Threshold Estimation”, *Econometrica*, **68**, 575–603.
- HARVEY, C. and SIDDIQUE, A. (1999), “Autoregressive Conditional Skewness”, *Journal of Financial and Quantitative Analysis*, **34**, 465–487.
- HARVEY, C. and SIDDIQUE, A. (2000), “Conditional Skewness in Asset Pricing Tests”, *Journal of Finance*, **55**, 1263–1296.
- HAUSMAN, J. (1978), “Specification Tests in Econometrics”, *Econometrica*, **46**, 1251–1271.
- HINICH, M. (1982), “Testing for Gaussianity and Linearity of a Stationary Time Series”, *Journal of Time Series Analysis*, **3**, 169–176.
- HINICH, M. and PATTERSON, D. (1993), “A New Diagnostic Test of Model Inadequacy Which Uses the Martingale Difference Criterion”, *Journal of Time Series Analysis*, **13**, 233–252.
- HJELLVIK, V. and TJØSTHEIM, D. (1996), “Nonparametric Statistics for Testing of Linearity and Serial Independence”, *Journal of Nonparametric Statistics*, **6**, 223–251.
- HONG, Y. (1999), “Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach”, *Journal of the American Statistical Association*, **84**, 1201–1220.
- HONG, Y. and LEE, T.-H. (2003), “Diagnostic Checking for Adequacy of Nonlinear Time Series Models”, *Econometric Theory*, **19**, 1065–1121.
- HOROWITZ, J. (2003), “Bootstrap Methods for Markov Processes”, *Econometrica*, **71**, 1049–1082.
- HSIEH, D. (1989), “Testing for Nonlinear Dependence in Daily Foreign Exchange Rates”, *Journal of Business*, **62**, 339–368.
- JIANG, G. J. and KNIGHT, J. L. (2002), “Estimation of Continuous Time Processes via the Empirical Characteristic Function”, *Journal of Business and Economic Statistics*, **20**, 198–212.
- JONDEAU, E. and ROCKINGER, M. (2003), “Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements”, *Journal of Economic Dynamics and Control*, **27**, 1699–1737.
- KEENAN, D. M. (1985), “A Tukey Non-Additivity-Type Test for Time Series Nonlinearity”, *Biometrika*, **72**, 39–44.
- KIM, W. and LINTON, O. (2003), “A Local Instrumental Variable Estimation Method for Generalized Additive Volatility Models” (Working Paper, Institute for Statistics and Econometrics, Humboldt University of Berlin, and Department of Economics, London School of Economics).
- KNIGHT, J. L. and YU, J. (2002), “Empirical Characteristic Function in Time Series Estimation”, *Econometric Theory*, **18**, 691–721.
- LEE, T.-H., WHITE, H. and GRANGER, C. W. J. (1993), “Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests”, *Journal of Econometrics*, **56**, 269–290.
- LI, Q. (1999), “Consistent Specification Tests for Time Series Econometric Models”, *Journal of Econometrics*, **92**, 101–147.
- LO, A. and MACKINLAY, A. C. (1988), “Stock Prices do not Follow Random Walks: Evidence from a Simple Specification Test”, *Review of Financial Studies*, **1**, 41–66.

- LOBATO, I. (2002), "A Consistent Test for the Martingale Difference Hypothesis", *Econometric Reviews* (forthcoming).
- LUUKKONEN, R., SAIKKONEN, P. and TERÄSVIRTA, T. (1988a), "Testing Linearity Against Smooth Transition Autoregressive Models", *Biometrika*, **75**, 491–499.
- LUUKKONEN, R., SAIKKONEN, P. and TERÄSVIRTA, T. (1988b), "Testing Linearity in Univariate Time Series Models", *Scandinavian Journal of Statistics*, **15**, 161–75.
- MARSRY, E. and TJØSTHEIM, D. (1997), "Additive Nonlinear ARX Time Series and Projection Estimates", *Econometric Theory*, **13**, 214–252.
- MEESE, R. and ROSE, A. (1991), "An Empirical Assessment of Non-Linearities in Models of Exchange Rate Determination", *Review of Economic Studies*, **58**, 603–619.
- PARK, J. and WHANG, Y. J. (2003), "Testing for the Martingale Hypothesis" (Working Paper, Department of Economics, Rice University and Department of Economics, Korea University).
- PINKSE, Y. (1998), "Consistent Nonparametric Testing for Serial Independence", *Journal of Econometrics*, **84**, 205–231.
- POTTER, S. (1995), "A Nonlinear Approach to U.S. GNP", *Journal of Applied Econometrics*, **10**, 109–125.
- PRIESTLEY, M. (1981) *Spectral Analysis and Time Series* (London: Academic Press).
- PRIESTLEY, M. (1988) *Non-Linear and Non-Stationary Time Series Analysis* (London: Academic Press).
- SAIKKONEN, P. and LUUKKONEN, R. (1988), "Lagrange Multiplier Tests for Testing Non-Linearities in Time Series Models", *Scandinavian Journal of Statistics*, **15**, 55–68.
- SARGENT, T. and LJUNGQVIST, L. (2002) *Recursive Macroeconomic Theory* (Cambridge, MA: MIT Press).
- SINGLETON, K. (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function", *Journal of Econometrics*, **102**, 111–141.
- STINCHCOMBE, M. and WHITE, H. (1998), "Consistent Specification Testing with Nuisance Parameters Present only Under the Alternative", *Econometric Theory*, **14**, 295–324.
- SUBBA RAO, T. and GABR, M. (1980), "A Test for Linearity of Stationary Time Series", *Journal of Time Series Analysis*, **1**, 145–158.
- TERÄSVIRTA, T. (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models", *Journal of the American Statistical Association*, **89**, 208–218.
- TERÄSVIRTA, T., LIN, C. J. and GRANGER, C. W. J. (1994a), "Power of the Neural Network Linearity Test", *Journal of Time Series Analysis*, **14**, 209–220.
- TERÄSVIRTA, T., TJØSTHEIM, D. and GRANGER, C. W. J. (1994b), "Aspects of Modelling Nonlinear Time Series", in R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics*, Vol. IV (North-Holland, Elsevier Science) 2917–2957.
- TIAO, G. and TSAY, R. S. (1994), "Some Advances in Nonlinear and Adaptive Modelling in Time Series", *Journal of Forecasting*, **13**, 109–131.
- TJØSTHEIM, D. (1994), "Non-Linear Time Series: A Selective Review", *Scandinavian Journal of Statistics*, **21**, 97–130.
- TONG, H. (1990) *Nonlinear Time Series: A Dynamic System Approach* (Oxford: Clarendon Press).
- TSAY, R. S. (1986), "Nonlinearity Tests for Time Series", *Biometrika*, **73**, 461–466.
- TSAY, R. S. (1989), "Testing and Modelling Threshold Autoregressive Processes", *Journal of the American Statistical Association*, **84**, 231–240.
- TSAY, R. S. (2001), "Nonlinear Time Series Models: Testing and Applications", in D. Peña, G. Tiao and R. S. Tsay (eds.) *A Course in Time Series Analysis* (New York: Wiley) 267–285.
- WHANG, Y.-J. (1998), "A Test of Autocorrelation in the Presence of Heteroscedasticity of Unknown Form", *Econometric Theory*, **14**, 87–122.
- WHITE, H. (1989), "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks", in *Proceeding of the International Joint Conference on Neural Networks* (New York: IEEE Press) 451–455.
- WOOLDRIDGE, J. (1990), "A Unified Approach to Robust, Regression-Based Specification Tests", *Econometric Theory*, **6**, 17–43.