Econometric Theory, **23**, 2007, 106–154. Printed in the United States of America.
DOI: 10.1017/S0266466607070053

AN IMPROVED GENERALIZED SPECTRAL TEST FOR CONDITIONAL MEAN MODELS IN TIME SERIES WITH CONDITIONAL HETEROSKEDASTICITY OF UNKNOWN FORM

Yongmiao Hong
Cornell University
and
Xiamen University

Yoon-Jin Lee
Indiana University

Dynamic economic theories usually have implications on and only on the conditional mean dynamics of economic processes. Using a generalized spectral derivative approach, Hong and Lee (2005, *Review of Economic Studies* 72, 499–541) recently proposed a new class of omnibus nonparametric specification tests for linear and nonlinear time series conditional mean models, where the dimension of the conditioning information set may be infinite. The tests can detect a wide range of model misspecifications in mean while being robust to conditional heteroskedasticity and time-varying higher order moments of unknown form. They enjoy an asymptotic "nuisance parameter–free" property in the sense that parameter estimation uncertainty has no impact on the asymptotic $N(0,1)$ distribution of the test statistics. As a result, only the estimated residuals from the null parametric model are needed to implement the tests, and no specific estimation is required.

Although parameter estimation uncertainty has no impact on the asymptotic distribution of the tests, it may have significant impact on the finite-sample distribution, and such an impact may become more substantial as the number of estimated parameters increases. In this paper, we adopt the Wooldridge (1990, *Econometric Theory* 6, 17–43) device for parametric *m*-tests to the Hong and Lee (2005) nonparametric tests to reduce the impact of parameter estimation uncer-

We thank Bruce Hansen (the co-editor) and three referees for insightful and helpful comments and suggestions. We also are grateful for financial support from the National Science Foundation Grant SES–0111769 and the Cheung Kong Visiting Scholarship by the Chinese Ministry of Education and Xiamen University, China. Address correspondence to Yongmiao Hong, Department of Economics and Department of Statistical Science, Uris Hall, Cornell University, Ithaca, NY 14850, USA, or Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, Fujian, China; e-mail: yh20@cornell.edu. Yoon-Jin Lee, Department of Economics, Indiana University, Wylie Hall 207, Bloomington, IN 47405, USA; e-mail: lee243@indiana.edu.

  0266-4666/07 $12.00

tainty. Asymptotic size and power properties of the modified tests are investigated, and simulation studies show that the modified tests generally have better sizes in finite samples and are robust to parameter estimation uncertainty. In the meantime, the size improvement does not cause loss of power against a wide range of alternatives when using the empirical critical values for the tests. These results suggest that the modified generalized spectral derivative tests can be a useful tool in time series conditional mean modeling.

1. INTRODUCTION

Most dynamic economic theories, such as the efficient market hypothesis, the expectations hypothesis, consumption and tax smoothing, dynamic asset pricing, and more generally rational expectations, have implications on and only on the conditional mean dynamics of underlying economic variables (e.g., Cochrane, 2001; Sargent and Ljungqvist, 2002; Adda and Cooper, 2003). For example, dynamic asset pricing implies that the expectation of the pricing error given the information available to economic agents is zero for all assets. Although economic theory may suggest a nonlinear relationship for the conditional mean dynamics, it does not give a concrete functional form. Various models used in practice can be, at best, viewed as approximations to the underlying conditional mean dynamics. It is important to check conditional mean specification, because misspecification in mean can lead to misleading conclusions and suboptimal point forecasts. Indeed, specification testing for dynamic conditional mean models has become an integral part of the modern time series econometric model building practice (e.g., White, 1987; Wooldridge, 1990a, 1990b, 1991).

Hong and Lee (2005) recently proposed a class of generally applicable omnibus nonparametric tests for possibly nonlinear time series conditional mean models, without requiring any prior knowledge of possible alternatives. They used a suitable partial derivative of the generalized spectrum that focuses on the conditional mean dynamics. The generalized spectrum was first proposed in Hong (1999) as a new frequency domain analytic tool for nonlinear time series. It can capture both linear and nonlinear serial dependence and enjoys the nice features of spectral analysis. In particular, it incorporates information on serial dependence at all lags and can characterize cyclical dynamics caused by linear or nonlinear serial dependence. As a result, the Hong and Lee (2005) tests can detect a wide variety of conditional mean misspecifications in both functional form and lag structure. This differs from existing tests for time series conditional mean models, which assume a fixed lag order and focus on functional form specification. One important feature of time series modeling is that the conditioning information set usually contains an infinite number of lags (i.e., the entire past history), unless a Markovian assumption holds. Hong and Lee (2005) checked a large number of lags without suffering from the "curse of dimensionality." Because they compared a nonparametric generalized spectral derivative estimator with a restricted counterpart implied by correct conditional mean specification, their tests can be viewed as a generalization of the Hausman (1978) methodology from a parametric context to a nonparametric time series context. Unlike Hausman (1978), however, Hong and Lee (2005) did not require that parameter estimators be asymptotically efficient under the null hypothesis.

Dynamic economic theory, although having implications on the conditional mean dynamics of an underlying economic process, is usually silent about its higher order conditional moment dynamics. There is a growing consensus among economists that the volatilities of most economic and financial time series are time-varying (e.g., Wooldridge, 1990a; Granger and Teräsvirta, 1993). Volatility clustering is a rule rather than an exception for most economic and financial time series. Moreover, recent studies (e.g., Gallant, Hsieh, and Tauchen, 1991; Hansen, 1994; Harvey and Siddique, 1999, 2000; Jondeau and Rockinger, 2003) documented time-varying conditional skewness and kurtosis of economic and financial time series. Time-varying higher order moments may be caused by (e.g.) correlated jumps or large sudden changes that occur occasionally. It is important to develop tests of the conditional mean models that are robust not only to conditional heteroskedasticity but also to time-varying higher order moments. Failure to accommodate these features will lead to distorted sizes (i.e., Type I errors) for the tests. The Hong and Lee (2005) tests are robust to conditional heteroskedasticity and all higher order conditional moments of unknown form. Thus, any conditional mean model can be subjected to testing without reestimating the model.

Hong and Lee (2005) only required estimation of the null conditional mean model, and parameter estimation uncertainty has no impact on the asymptotic distribution of their tests. Intuitively, parameter estimation uncertainty has an impact on the finite-sample distribution of the tests, and the degree of the impact depends on the number of estimated parameters associated with endogenous variables. For a parametric model, the number of estimated parameters is finite and fixed as the sample size $T \to \infty$. As a result, the impact is at most an adjustment of a finite number of degrees of freedom. When the number of lags employed in the generalized spectral derivative tests is large (i.e., grows to infinity with T), the adjustment of a finite number of degrees of freedom becomes asymptotically negligible.

The asymptotic "nuisance parameter–free" property simplifies the implementation of the tests. Only the estimated model residuals are needed to implement the tests, and no specific estimation method is required. However, the sample sizes of most low-frequency economic and financial time series data are not large, whereas econometric time series models may contain a relatively large number of estimated parameters. Thus, the impact of parameter estimation uncertainty may not be trivial in small and finite samples. Indeed, our simulation studies show that the empirical sizes of the Hong and Lee (2005) tests deteriorate as the number of estimated autoregressive parameters increases.

In particular, they tend to underreject, apparently because parameter estimation tends to make the estimated residuals look more like a martingale difference sequence (m.d.s.). This can be troublesome in practice, because underrejection makes it more difficult to detect neglected dynamic structure in mean in finite samples.

To deal with the underrejection problem, one could use a bootstrap procedure, which can take into account the impact of parameter estimation uncertainty. In the present context, naive bootstraps cannot be used, because under correct specification of the conditional mean model, the regression error is an m.d.s., which is not necessarily an independent and identically distributed (i.i.d.) sequence. For a non-i.i.d. sequence, the bootstrap can be complicated, because higher order serial dependence has to be preserved (Horowitz, 2003; Gonçalves and Kilian, 2004). In this paper, we propose a substantive modification to the Hong and Lee (2005) tests that can reduce the impact of parameter estimation uncertainty. This is achieved by using the Wooldridge (1990a) device, which is a convenient auxiliary regression that can effectively remove the impact of parameter estimation uncertainty of a parametric test statistic up to a higher order. By running an increasing sequence of auxiliary regressions, we can reduce the impact of parameter estimation uncertainty of the generalized spectral derivative tests. As a consequence, the finite-sample distribution of the tests is expected to be more robust to the number of estimated parameters.

As White (1994) pointed out, the Wooldridge (1990a) device generally renders a test unable to detect certain misspecification. This is also the case for our modified generalized spectral derivative tests. However, when the least squares estimator (or a Gaussian quasi–maximum likelihood estimator) is used, the modified generalized spectral derivative tests share the same consistency (i.e., asymptotic power one) property as the unmodified generalized spectral derivative tests, although the modified and unmodified test statistics, after being properly scaled, do not converge to the same probability limit under a fixed alternative. In this case, there is no asymptotic power loss for our modified tests, but their sizes have been significantly improved. This is confirmed in our simulation studies.

Section 2 introduces hypotheses of interest and modified generalized spectral derivative tests and also provides heuristics on how the Wooldridge (1990a) device can improve the asymptotic normal approximation of the generalized spectral derivative tests. Section 3 derives the asymptotic distribution of the modified tests, and Section 4 investigates their asymptotic power property under a general fixed alternative. In Section 5, we compare the finite-sample performances of the modified and unmodified generalized spectral derivative tests. Section 6 concludes. All proofs are collected in the Appendix. The GAUSS code for implementing our modified 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. GENERALIZED SPECTRAL DERIVATIVE TESTS

2.1. Hypotheses of Interest

Suppose $g(I_{t-1},\theta)$ is a parametric model for the conditional mean $E(Y_t|I_{t-1})$ of a stochastic time series process $\{Y_t\}$, where I_{t-1} is an information set at time $t-1$, which may contain lagged dependent variables $\{Y_{t-j}, j>0\}$ and current and lagged exogenous variables $\{Z_{t-j}, j\geq 0\}$, and $\theta\in\Theta$ is a finite-dimensional parameter. Examples of $g(I_{t-1},\theta)$ include linear time series regression (static or dynamic) models, autoregressive moving average (ARMA) models, ARMA with exogenous variables (ARMAX) models, regime-switching autoregressive models (Hamilton, 1989), parametric state-space models (Priestley, 1988), smooth transition autoregressive models (Teräsvirta, 1994), Poisson jump autoregressive models, and threshold autoregressive models with known thresholds (e.g., Potter, 1995).[1]

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

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

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

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

Conditional mean modeling has been a 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, most dynamic economic theories have implications on and only on the conditional mean dynamics of economic processes, as pointed out earlier.

2.2. Generalized Spectral Derivative Tests

In time series modeling, 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 the adequacy of the model $g(I_{t-1},\theta)$, due to the curse of dimensionality. Hong and Lee (2005) proposed a nonparametric test of $\mathbb{H}_0$ using a suitable partial derivative of the Hong (1999) generalized spectrum, which avoids the curse of dimensionality. Define the model error

$$\varepsilon_t(\theta)\equiv Y_t-g(I_{t-1},\theta), \qquad \theta\in\Theta. \tag{2.1}$$

Then $\mathbb{H}_0$ holds if and only if $E[\varepsilon_t(\theta_0)|I_{t-1}]=0$ *a.s.* for some $\theta_0\in\Theta$. The null hypothesis $\mathbb{H}_0$ thus implies $E[\varepsilon_t(\theta_0)|I_{t-1}^{\varepsilon}]=0$ *a.s.*, where $I_{t-1}^{\varepsilon}\equiv\{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0),\ldots\}$. This forms a basis for testing $\mathbb{H}_0$.

For notational economy, we put $\varepsilon_t\equiv\varepsilon_t(\theta_0)$, where $\theta_0=p\lim\hat{\theta}$ and $\hat{\theta}$ is a parameter estimator. 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 $j = 0, \pm 1, \ldots$. The basic idea of the generalized spectrum in Hong (1999) is to consider the spectrum of the transformed series $\{e^{iu\varepsilon_t}\}$. It is defined as

$$f(\omega,u,v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v) e^{-ij\omega}, \qquad \omega \in [-\pi,\pi], u,v \in \mathbb{R}, \tag{2.2}$$

where ω is the frequency and $\sigma_j(u,v) \equiv \text{cov}(e^{iu\varepsilon_t}, e^{iv\varepsilon_{t-|j|}})$ is the covariance function of the transformed series. 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 $j \neq 0$, including nonlinear serial dependence with zero autocorrelation. This is analogous to the higher order spectra (Brillinger, 1965; Brillinger and Rosenblatt, 1967a, 1967b). Unlike the higher order spectra, however, $f(\omega,u,v)$ does not require the existence of any moment of $\{\varepsilon_t\}$. Nevertheless, when $E(\varepsilon_t^2)$ exists, we can obtain the power spectrum as a partial derivative of $f(\omega,u,v)$ at $(u,v) = (0,0)$:

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

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 more difficult for nonlinear time series than for linear time series. For example, the bispectrum (e.g., Subba Rao and Gabr, 1984) has no physical (i.e., energy decomposition over frequencies) interpretation, unlike the power spectrum. This is also true of $f(\omega,u,v)$. However, the basic idea of characterizing cyclical dynamics still applies: $f(\omega,u,v)$ is useful when searching for linear or nonlinear cyclical movements. A strong cyclicity of data can be linked with a strong serial dependence in $\{\varepsilon_t\}$ 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. This can be seen from the Taylor series expansion of $f(\omega,\cdot,\cdot)$ around the origin (0,0):

$$f(\omega,u,v) = \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \frac{(iu)^m (iv)^l}{m!\,l!} \left[\frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \text{cov}(\varepsilon_t^m, \varepsilon_{t-|j|}^l) e^{-ij\omega} \right],$$

$$\omega \in [-\pi,\pi], u,v \in \mathbb{R},$$

assuming all moments of $\{\varepsilon_t\}$ exist. Now suppose an asset series is a white noise ($\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ for all $j \neq 0$) but has a stochastic cyclical dynamics in volatility clustering, which may be linked to business cycles (e.g., Schwert, 1989; Hamilton and Lin, 1996). Then the power spectrum will miss these volatility cycles, but $f(\omega,u,v)$ can effectively capture them. More generally, $f(\omega,u,v)$ can capture cyclical dynamics in the conditional distribution of $\{\varepsilon_t\}$, including those in volatility, skewness, and kurtosis.[2]

The generalized spectrum $f(\omega,u,v)$ itself is not suitable for testing $\mathbb{H}_0$, because it can capture serial dependence not only in mean but also in higher order moments.[3] An example is an autoregressive conditionally heteroskedastic (ARCH) process, which is an m.d.s. but can be captured by $f(\omega,u,v)$. 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 serial dependence in various moments. To focus on and only on serial dependence in mean, one can use the partial derivative

$$f^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\sigma_j^{(1,0)}(0,v)e^{-ij\omega}, \qquad \omega \in [-\pi,\pi], v \in \mathbb{R}, \tag{2.3}$$

where

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

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 regularity 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.*[4] Therefore, $\sigma_j^{(1,0)}(0,v)$, or equivalently $f^{(0,1,0)}(\omega,0,v)$, ideally suits for testing conditional mean dynamics. The autoregression function can capture linear and nonlinear serial dependence in mean, including 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 $\{z_t\} \sim i.i.d.(0,\sigma^2)$.[5]

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. Moreover, $\sup_{v\in\mathbb{R}}|\sigma_j^{(1,0)}(0,v)|$ can be viewed as an operational measure of the maximum mean correlation, $\max_{f(\cdot)}|\operatorname{corr}[\varepsilon_t, f(\varepsilon_{t-j})]|$, which was proposed by Granger and Teräsvirta (1993, p. 23) as a measure for nonlinearity in mean. Similarly, the generalized spectral derivative modulus

$$m(\omega) \equiv \sup_{v\in(-\infty,\infty)}|f^{(0,1,0)}(\omega,0,v)|$$

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 (e.g., ARCH-in-mean effect; see Engle, Lilien, and Robins, 1987).[6]

Because ε_t is not observed, we need to use an estimated model residual

$$\hat{\varepsilon}_t \equiv Y_t - g(I_{t-1}^{\dagger}, \hat{\theta}), \qquad t = 1, \ldots, T, \tag{2.4}$$

where $I_{t-1}^{\dagger}$ is the information set observed at time $t-1$ that may involve some assumed initial values.[7] Any $\sqrt{T}$-consistent parameter estimator $\hat{\theta}$ can be used. Examples of $\hat{\theta}$ are the conditional least squares and quasi–maximum likelihood estimators. With $\{\hat{\varepsilon}_t\}_{t=1}^T$, one can consistently estimate $f^{(0,1,0)}(\omega,0,v)$ by a smoothed kernel estimator

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

where

$$\hat{\sigma}_j^{(1,0)}(0,v) = \frac{1}{T-|j|} \sum_{t=|j|+1}^{T} i\hat{\varepsilon}_t \hat{\psi}_{t-|j|}(v), \tag{2.5}$$

$\hat{\psi}_{t-|j|}(v) = e^{iv\hat{\varepsilon}_{t-|j|}} - \hat{\varphi}_j(v)$, and $\hat{\varphi}_j(v) = (T-|j|)^{-1} \sum_{t=|j|+1}^{T} e^{iv\hat{\varepsilon}_{t-|j|}}$. Here, $p \equiv p(T)$ is a bandwidth, and $k:\mathbb{R} \to [-1,1]$ is a symmetric kernel. Examples of $k(\cdot)$ include the Bartlett, Daniell, Parzen, and quadratic spectral kernels (e.g., Priestley, 1981, p. 442). The factor $(1-|j|/T)^{1/2}$ is a finite-sample correction. It could be replaced by unity.

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

$$f_0^{(0,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi} \sigma_0^{(1,0)}(0,v), \qquad \omega \in [-\pi,\pi], u,v \in \mathbb{R},$$

which can be consistently estimated by

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

To test $\mathbb{H}_0$, Hong and Lee (2005) compared $\hat{f}^{(0,1,0)}(\omega,0,v)$ and $\hat{f}_0^{(0,1,0)}(\omega,0,v)$ via an L_2-norm. Their test statistic is

$$\hat{M}_1(p) \equiv \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\hat{\sigma}_j^{(1,0)}(0,v)|^2 \, dW(v) - \hat{C}_1(p) \right] \Big/ \sqrt{\hat{D}_1(p)}, \tag{2.6}$$

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

$$\hat{C}_1(p) = \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=j+1}^{T-1} \hat{\varepsilon}_t^2 \int |\hat{\psi}_{t-j}(v)|^2 \, dW(v), \quad \text{and}$$

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

Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. An example of $W(\cdot)$ is the $N(1,0)$ cumulative distribution function (c.d.f.), which is commonly used in the characteristic function literature. The factors $\hat{C}_1(p)$ and $\hat{D}_1(p)$ are the approximate mean and variance of the quadratic form $T\int\int_{-\pi}^{\pi} |\hat{f}^{(0,1,0)}(\omega,0,v) - \hat{f}_0^{(0,1,0)}(\omega,0,v)|^2\, d\omega dW(v)$. They have taken into account the impact of conditional heteroskedasticity and time-varying higher order conditional moments. As a result, $\hat{M}_1(p)$ is robust to conditional heteroskedasticity and time-varying higher order conditional moments of unknown form.[8] We note that both $\hat{C}_1(p)$ and $\hat{D}_1(p)$ grow to infinity at a rate of p as $p \to \infty$, $p/T \to 0$.

Hong and Lee (2005) showed that $\hat{M}_1(p) \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$. Because the parametric estimator $\hat{\theta}$ converges to θ_0 faster than the rate at which the nonparametric kernel estimator $\hat{f}^{(0,1,0)}(\omega,0,v)$ converges to $f^{(0,1,0)}(\omega,0,v)$, the asymptotic normal distribution of $\hat{M}_1(p)$ is solely determined by $\hat{f}^{(0,1,0)}(\omega,0,v)$. Attractively, $\hat{\theta}$ can be any $\sqrt{T}$-consistent estimator of θ_0, and parameter estimation uncertainty has no impact on the asymptotic distribution of $\hat{M}_1(p)$. In other words, the asymptotic distribution of $\hat{M}_1(p)$ is unchanged when $\hat{\theta}$ is replaced by its probability limit θ_0. This results in significant simplification of some otherwise difficult contexts. In particular, only estimated model residuals are needed to compute the test statistics.

However, the convenient asymptotic nuisance parameter–free property is not without cost for $\hat{M}_1(p)$. Although parameter estimation uncertainty has no impact on the asymptotic distribution of $\hat{M}_1(p)$, it has an impact on the finite-sample distribution. Because $g(I_{t-1},\theta)$ is a parametric model, $\hat{\theta}$ can result in at most an adjustment of a finite number of degrees of freedom to the distribution of $\hat{M}_1(p)$. When the lag order $p \to \infty$ as $T \to \infty$, the impact of $\hat{\theta}$ becomes negligible when normalized by the standard deviation estimator $\hat{D}_1(p)^{1/2}$, which grows to infinity at the rate of $p^{1/2}$. However, asymptotic analysis reveals that the asymptotically negligible higher order terms in $\hat{M}_1(p)$ that are associated with parameter estimation uncertainty vanish to zero in probability rather slowly, as will be seen subsequently. Therefore, $\hat{\theta}$ may significantly distort the size of $\hat{M}_1(p)$ in small and finite samples. This is particularly the case when there are relatively many parameters but the sample size T is not large, as is typically encountered for macroeconomic time series data and low-frequency financial time series data.

2.3. Wooldridge's Device

In a series of important works, Wooldridge (1990a, 1990b, 1991) proposed a unified approach to robust, regression-based parametric specification tests for possibly dynamic time series models, including time series conditional mean models. Wooldridge (1990a) considered the null hypothesis

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

where $\phi_t(\theta)$ is a measurable, possibly vector-valued function. In the present context, $\phi_t(\theta) = i\varepsilon_t(\theta)$ in (2.1). Wooldridge (1990a) used a test function $\Lambda_t(\theta) \in I_{t-1}$ and checked if $E[\Lambda_t(\theta_0)\phi_t(\theta_0)] = 0$ by using the sample moment

$$\hat{m} \equiv \frac{1}{T}\sum_{t=1}^{T}\hat{\Lambda}_t\hat{\phi}_t,$$

where $\hat{\Lambda}_t = \Lambda_t(\hat{\theta})$, $\hat{\phi}_t = \phi_t(\hat{\theta})$, and $\hat{\theta}$ is a $\sqrt{T}$-consistent estimator of θ_0. Straightforward algebra shows that

$$\sqrt{T}\hat{m} = T^{-1/2}\sum_{t=1}^{T}[\Lambda_t(\theta_0)\phi_t(\theta_0) + \Lambda_t(\theta_0)\Phi_t(\theta_0)(\hat{\theta} - \theta_0)] + O_P(T^{-1/2}),$$

where $\Phi_t(\theta_0) \equiv E[(\partial/\partial\theta)\phi_t(\theta_0)|I_{t-1}]$. Thus, the asymptotic distribution of $\sqrt{T}\hat{m}$ is jointly determined by $T^{-1/2}\sum_{t=1}^{T}\Lambda_t(\theta_0)\phi_t(\theta_0)$ and $\sqrt{T}(\hat{\theta} - \theta_0)$, unless $\Phi_t(\theta_0) = 0$ under the null hypothesis.

To remove the impact of parameter estimation uncertainty of $\hat{\theta}$ on the asymptotic distribution of $\sqrt{T}\hat{m}$, Wooldridge (1990a) first purged from $\hat{\Lambda}_t$ its linear projection onto $\hat{\Phi}_t$, a consistent estimator for $\Phi_t(\theta_0)$, and then considered the modified test statistic

$$\hat{m}_d \equiv \frac{1}{T}\sum_{t=1}^{T}(\hat{\Lambda}_t - \hat{\Phi}_t'\hat{\beta})\hat{\phi}_t,$$

where $\hat{\beta}$ is the ordinary least squares (OLS) estimator of regressing $\hat{\Lambda}_t$ on $\hat{\Phi}_t$. It can be shown that for any $\sqrt{T}$-consistent estimator $\hat{\theta}$,

$$\sqrt{T}\hat{m}_d = T^{-1/2}\sum_{t=1}^{T}[\Lambda_t(\theta_0) - \Phi_t(\theta_0)'\beta]\phi_t(\theta_0) + O_P(T^{-1/2}),$$

where $\beta \equiv p\lim\hat{\beta}$. Thus, the asymptotic distribution of $\sqrt{T}\hat{m}_d$ is robust to parameter estimation uncertainty because it is not affected by any $\sqrt{T}$-consistent estimator $\hat{\theta}$ up to $O_P(T^{-1/2})$. An asymptotic χ^2 test can be obtained by forming a suitable quadratic form in $\sqrt{T}\hat{m}_d$. We note that the Wooldridge (1990a) device does not imply that $\sqrt{T}\hat{m}_d$ has a better asymptotic approximation than $\sqrt{T}\hat{m}$ in finite samples or vice versa. However, it nicely generates a new set of moment conditions $\sqrt{T}\hat{m}_d$ that is robust to parameter estimation uncertainty up to $O_P(T^{-1/2})$. Consequently, its asymptotic distribution does not depend on any $\sqrt{T}$-consistent estimator $\hat{\theta}$. This makes the test based on $\sqrt{T}\hat{m}_d$ rather convenient in practice.

Although the Wooldridge (1990a) device may not deliver a better asymptotic distribution approximation for $\sqrt{T}\hat{m}_d$, it ideally suits our purpose of improving the finite-sample performance of the generalized spectral derivative test $\hat{M}_1(p)$ in (2.6), because with a new set of moment conditions, it can make the asymptotically negligible higher order terms in $\hat{M}_1(p)$ that are associated with $\hat{\theta}$ vanish faster to 0. Subsequently, we first describe how the Wooldridge (1990a) device can be adapted to $\hat{M}_1(p)$ and then explain why it can improve the asymptotic normal approximation for $\hat{M}_1(p)$.

Although $\hat{M}_1(p)$ is more complicated than the Wooldridge (1990a) test statistic, the Wooldridge (1990a) device can be applied to each generalized covariance derivative $\hat{\sigma}_j^{(1,0)}(0,v)$, which has a similar structure to $\hat{m}$, with $\hat{\phi}_t = i\hat{\varepsilon}_t$ and $\hat{\Lambda}_t = \hat{\psi}_{t-|j|}(v)$. Based on this observation, we introduce a modified generalized covariance

$$\hat{\gamma}_j^{(1,0)}(0,v) = (T-|j|)^{-1}\sum_{t=|j|+1}^{T} i\hat{\varepsilon}_t\hat{h}_{t-|j|}(v), \qquad j=0,\pm1,\dots,\pm(T-1), \quad \textbf{(2.7)}$$

where $\hat{h}_{t-|j|}(v) = \hat{\psi}_{t-|j|}(v) - \hat{G}_t'\hat{\beta}_{|j|}(v)$, $\hat{G}_t = (\partial/\partial\theta)g(I_{t-1}^{\dagger},\hat{\theta})$, and

$$\hat{\beta}_j(v) = \left(\sum_{t=1}^{T}\hat{G}_t\hat{G}_t'\right)^{-1}\sum_{t=|j|+1}^{T}\hat{G}_t\hat{\psi}_{t-|j|}(v). \quad \textbf{(2.8)}$$

The modified generalized spectral derivative estimators can then be defined as follows:

$$\hat{S}^{(0,1,0)}(\omega,0,v) = \frac{1}{2\pi}\sum_{j=1-T}^{T-1}k(j/p)(1-|j|/T)^{1/2}\hat{\gamma}_j^{(1,0)}(0,v)e^{-ij\omega},$$

$$\omega\in[-\pi,\pi], \qquad v\in\mathbb{R}, \quad \textbf{(2.9)}$$

and

$$\hat{S}_0^{(0,1,0)}(\omega,0,v) = \frac{1}{2\pi}\hat{\gamma}_0^{(1,0)}(0,v), \qquad \omega\in[-\pi,\pi], \qquad v\in\mathbb{R}.$$

Comparing $\hat{S}^{(0,1,0)}(\omega,0,v)$ and $\hat{S}_0^{(0,1,0)}(\omega,0,v)$ via an L_2-norm, we obtain the modified test statistic

$$\hat{M}_1^d(p) = \left[\sum_{j=1}^{T-1}k^2(j/p)(T-j)\int|\hat{\gamma}_j^{(1,0)}(0,v)|^2\,dW(v) - \hat{C}_1^d(p)\right]\Big/\sqrt{\hat{D}_1^d(p)}, \quad \textbf{(2.10)}$$

where the centering and scaling factors

$$\hat{C}_1^d(p) = \sum_{j=1}^{T-1} k^2(j/p) \frac{1}{T-j} \sum_{t=j+1}^{T} \hat{\varepsilon}_t^2 \int |\hat{h}_{t-j}(v)|^2 \, dW(v) \quad \text{and}$$

$$\hat{D}_1^d(p) = 2 \sum_{j=1}^{T-2} \sum_{l=1}^{T-2} k^2(j/p) k^2(l/p)$$
$$\times \iint \left| \frac{1}{T-\max(j,l)} \sum_{t=j+1}^{T} \hat{\varepsilon}_t^2 \hat{h}_{t-j}(v) \hat{h}_{t-l}(v') \right|^2 dW(v) \, dW(v').$$

To gain insight into why the Wooldridge (1990a) device can improve the finite-sample performance of $\hat{M}_1(p)$, we now provide heuristics. Put $\psi_t(v) = e^{iv\varepsilon_t} - \varphi(v)$, $G_t = (\partial/\partial\theta) g(I_{t-1}, \theta_0)$, $\xi_{t-j}(v) = ivG_{t-j} e^{iv\varepsilon_{t-j}}$, and $\eta_j(v) = E[G_t \psi_{t-j}(v)]$ for $j > 0$ and let $\tilde{\sigma}_j^{(1,0)}(0,v)$ be defined in the same way as $\hat{\sigma}_j^{(1,0)}(0,v)$ with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. Then, by a Taylor series expansion around θ_0, we have for each $j > 0$,

$$\begin{aligned}
\hat{\sigma}_j^{(1,0)}&(0,v) \\
&= \tilde{\sigma}_j^{(1,0)}(0,v) - (\hat{\theta} - \theta_0)' \frac{1}{T-j} \sum_{t=j+1}^{T} iG_t \psi_{t-j}(v) \\
&\quad - (\hat{\theta} - \theta_0)' \frac{1}{T-j} \sum_{t=j+1}^{T} i\varepsilon_t \xi_{t-j}(v) + O_P[(T-j)^{-1}] \\
&= \tilde{\sigma}_j^{(1,0)}(0,v) - (\hat{\theta} - \theta_0)' i\eta_j(v) \\
&\quad - (\hat{\theta} - \theta_0)' \frac{1}{T-j} \sum_{t=j+1}^{T} i[G_t \psi_{t-j}(v) - \eta_j(v)] \\
&\quad - (\hat{\theta} - \theta_0)' \frac{1}{T-j} \sum_{t=j+1}^{T} i\varepsilon_t \xi_{t-j}(v) + O_P[(T-j)^{-1}] \\
&= \tilde{\sigma}_j^{(1,0)}(0,v) - (\hat{\theta} - \theta_0)' i\eta_j(v) + O_P[(T-j)^{-1}],
\end{aligned} \tag{2.11}$$

where the remainder $O_P[(T-j)^{-1}]$ term in the first equality comes from the higher order terms of the Taylor series expansion, including the effect of replacing the information set $I_{t-1}^{\dagger}$ with I_{t-1}. The last equality follows from $\hat{\theta} - \theta_0 = O_P(T^{-1/2})$, Chebyshev's inequality, $\sup_{v \in \mathbb{R}} E\|\sum_{t=j+1}^{T} \varepsilon_t \xi_{t-j}(v)\|^2 \le C(T-j)$ under the m.d.s. property of $\{\varepsilon_t\}$ as implied by $\mathbb{H}_0$, and $E\|\sum_{t=j+1}^{T} [G_t \psi_{t-j}(v) - \eta_j(v)]\|^2 \le C(T-j)$ under suitable mixing conditions on the time series process $\{\varepsilon_t, G_t'\}'$.

For static conditional mean models where $g(I_{t-1}, \theta_0)$ is only a function of strictly exogenous variables independent of innovations $\{\varepsilon_t\}$, we have $\eta_j(v) = 0$ for all $j > 0$. For dynamic conditional mean models where $g(I_{t-1}, \theta_0)$ is a function of lagged dependent variables and/or lagged innovations, $\eta_j(v)$ is generally nonzero at least for some $j > 0$. To examine the impact of $\eta_j(v)$ on $\hat{M}_1(p)$, we use (2.11) and obtain

$$\begin{aligned}
&\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\hat{\sigma}_j^{(1,0)}(0,v)|^2 \, dW(v) \\
&\quad = \sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \, dW(v) \\
&\qquad + (\hat{\theta}-\theta_0)' \left[\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int \eta_j(v)\eta_j^*(v)' \, dW(v) \right] (\hat{\theta}-\theta_0) \\
&\qquad - 2(\hat{\theta}-\theta_0)' \sum_{j=1}^{T-1} k^2(j/p)(T-j) \mathrm{Re} \int i\eta_j(v) \tilde{\sigma}_j^{(1,0)}(0,v)^* \, dW(v) \\
&\qquad + O_P(p/T^{1/2}) \\
&\quad = \sum_{j=1}^{T-1} k^2(j/p) T_j \int |\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \, dW(v) + O_P(1) + O_P(1) + O_P(p/T^{1/2}),
\end{aligned} \tag{2.12}$$

where the second term in the first equality is $O_P(1)$ given $\sqrt{T}(\hat{\theta}-\theta_0) = O_P(1)$ and

$$\sum_{j=1}^{T-1} k^2(j/p)(1-j/T) \int \eta_j(v)\eta_j^*(v)' \, dW(v) \to \sum_{j=1}^{\infty} \int \eta_j(v)\eta_j^*(v)' \, dW(v) = O(1)$$

by the dominated convergence theorem, where the limit is nonzero when $g(I_{t-1}, \theta_0)$ is a dynamic conditional mean model. Similarly, the third term in the first equality is also $O_P(1)$ given $E|\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \le C(T-j)^{-1}$ and Chebyshev's inequality.

Using analogous reasoning, we can also obtain that, for the mean and variance estimators $\hat{C}_1(p)$ and $\hat{D}_1(p)$ in $\hat{M}_1(p)$,

$$\hat{C}_1(p) = \tilde{C}_1(p) + O_P(p/T^{1/2}), \quad \text{and} \quad \hat{D}_1(p) = \tilde{D}_1(p) + O_P(p/T^{1/2}),$$

where $\tilde{C}_1(p)$ and $\tilde{D}_1(p)$ are defined in the same way as $\hat{C}_1(p)$ and $\hat{D}_1(p)$ with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. Recalling that $\hat{D}_1(p)$ grows at a rate of p, we then obtain

$$\hat{M}_1(p) = \widetilde{M}_1(p) + O_P(p^{-1/2}) + O_P(p^{1/2}/T^{1/2}), \tag{2.13}$$

where $\widetilde{M}_1(p)$ is an infeasible test statistic that is defined in the same way as $\hat{M}_1(p)$ with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. In (2.13), the $O_P(p^{-1/2})$ term arises because of the effect of $\hat{\theta}$ and the fact that $\hat{D}_1(p)^{1/2} \propto p^{1/2}$ in probability. This term vanishes to 0 in probability given $p \to \infty$. On the other hand, the

$O_P(p^{1/2}/T^{1/2})$ term also vanishes to 0 in probability given $p/T \to 0$. As a result, the asymptotic distribution of $\hat{M}_1(p)$ is determined by the infeasible test statistic $\widetilde{M}_1(p)$, which is evaluated at θ_0 and is asymptotically $N(0,1)$ under $\mathbb{H}_0$.

In practice, the lag order p usually grows to infinity at a slow rate. For example, when the Bartlett and Parzen kernels are used, the optimal rates for p in terms of the mean squared error criterion are $p \propto T^{1/3}$ and $p \propto T^{1/5}$, respectively. In these cases we have $p^{-1/2} \propto T^{-1/6}$ and $p^{-1/2} \propto T^{-1/10}$, respectively. The slow convergence of the $O_P(p^{-1/2})$ term implies that the asymptotic normal approximation for $\hat{M}_1(p)$ may be inadequate in finite samples and may become worse when one has to estimate more parameters.

The slowly vanishing $O_P(p^{-1/2})$ term in $\hat{M}_1(p)$ is thus troublesome in finite samples. The ability to remove it is highly desirable. This will improve the asymptotic $N(0,1)$ approximation for $\hat{M}_1(p)$ in finite samples. As we illustrate subsequently, the Wooldridge (1990a) device ideally suits this purpose, although it does not necessarily improve the size performance of the Wooldridge (1990a) modified parametric m-tests in finite samples. Let $\tilde{\gamma}_j^{(1,0)}(0,v)$ be defined in the same way as $\hat{\gamma}_j^{(1,0)}(0,v)$ with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. Then, by taking a Taylor series expansion and using reasoning analogous to that of (2.11), we have for $j > 0$,

$$\begin{aligned}\hat{\gamma}_j^{(1,0)}(0,v) &= \tilde{\gamma}_j^{(1,0)}(0,v) - (\hat{\theta} - \theta_0)' \frac{1}{T-j} \sum_{j=1}^{T-1} iG_t h_{t-j}(v) \\ &\quad - (\hat{\theta} - \theta_0)' \frac{1}{T-j} \sum_{t=j+1}^{T} i\varepsilon_t \delta_{t-j}(v) + O_P[(T-j)^{-1}] \\ &= \tilde{\gamma}_j^{(1,0)}(0,v) - (\hat{\theta} - \theta_0)' i\varsigma_j(v) + O_P[(T-j)^{-1}], \end{aligned} \qquad \textbf{(2.14)}$$

where $h_{t-j}(v) = \psi_{t-j}(v) - G_t'\beta_j(v)$, $\beta_j(v) = [E(G_tG_t')]^{-1}E[G_t\psi_{t-j}(v)]$, $\delta_{t-j}(v) = \xi_{t-j}(v) - G_t'\tau_j(v)$, $\tau_j(v) = [E(G_tG_t')]^{-1}E[G_t\xi_{t-j}(v)]$, and $\varsigma_j(v) = E[G_t h_{t-j}(v)]$. Unlike the function $\eta_j(v) = E[G_t\psi_{t-j}(v)]$ that is associated with $\hat{\sigma}_j^{(1,0)}(0,v)$ in (2.11), we always have $\varsigma_j(v) = 0$ for all $j > 0$ in (2.14), no matter whether $g(I_{t-1}, \theta_0)$ is a static or dynamic conditional mean model. It follows that $\hat{\gamma}_j^{(1,0)}(0,v) = \tilde{\gamma}_j^{(1,0)}(0,v) + O_P[(T-j)^{-1}]$, and consequently,

$$\begin{aligned}&\sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\hat{\gamma}_j^{(1,0)}(0,v)|^2 \, dW(v) \\ &\quad = \sum_{j=1}^{T-1} k^2(j/p)(T-j) \int |\tilde{\gamma}_j^{(1,0)}(0,v)|^2 \, dW(v) + O_P(p/T^{1/2}).\end{aligned}$$

Therefore, we have

$$\hat{M}_1^d(p) = \widetilde{M}_1^d(p) + O_P(p^{1/2}/T^{1/2}), \qquad \textbf{(2.15)}$$

where $\widetilde{M}_1^d(p)$ is an infeasible test statistic that is defined in the same way as $\hat{M}_1^d(p)$ with $\{\varepsilon_t\}_{t=1}^T$ replacing $\{\hat{\varepsilon}_t\}_{t=1}^T$. Under $\mathbb{H}_0$, we have $\widetilde{M}_1^d(p) \xrightarrow{d} N(0,1)$. The Wooldridge (1990a) device does not imply that $\widetilde{M}_1^d(p)$ has a better normal approximation than $\widetilde{M}_1(p)$ in finite samples or vice versa.[9] However, the $O_P(p^{-1/2})$ term in (2.13) now disappears in (2.15), and the $O_P(p^{1/2}/T^{1/2})$ term vanishes to 0 in probability much faster than $O_P(p^{-1/2})$. For example, when $p \propto T^{1/5}$, we have $p^{1/2}/T^{1/2} = T^{-2/5}$(compare $p^{-1/2} \propto T^{-1/10}$). We thus expect finite-sample improvement of the normal approximation for $\hat{M}_1^d(p)$, because its asymptotically negligible higher order terms vanish to 0 in probability faster than the higher order terms in $\hat{M}_1(p)$. We emphasize that the finite-sample improvement is achieved by combining Wooldridge's device and the nonparametric testing approach. As pointed out earlier, Wooldridge's device alone does not necessarily improve the finite-sample performance. Intuitively, for each $\hat{\sigma}_j^{(1,0)}(0,v)$, there is an impact of parameter estimation uncertainty. In particular, a Taylor series expansion of $(T-j)^{1/2}\hat{\sigma}_j^{(1,0)}(0,v)$ around θ_0 reveals that replacing $\hat{\theta}$ for θ_0 affects the asymptotic distribution, although the cumulative effect of replacing $\hat{\theta}$ for θ_0 becomes asymptotically negligible when we employ an increasing number of lags, making $\hat{M}_1(p)$ robust to parameter estimation uncertainty (this differs from the parametric m-tests based on $\sqrt{T}\hat{m}$). In contrast, a Taylor series expansion of $(T-j)^{1/2}\hat{\gamma}_j^{(1,0)}(0,v)$ around θ_0 reveals that the asymptotic distribution of $(T-j)^{1/2}\hat{\gamma}_j^{(1,0)}(0,v)$ is the same as that of $(T-j)^{1/2}\tilde{\gamma}_j^{(1,0)}(0,v)$. By applying the Wooldridge (1990) device to each $\hat{\sigma}_j^{(1,0)}(0,v)$ and using $\hat{\gamma}_j^{(1,0)}(0,v)$ instead, we can effectively reduce the impact of parameter estimation uncertainty to a higher order. Thus, we expect more robustness of $\hat{M}_1^d(p)$ to parameter estimation uncertainty in finite samples.

One important difference between the Wooldridge (1990a) test and $\hat{M}_1^d(p)$ is that Wooldridge (1990a) checks a fixed number of moment conditions whereas $\hat{M}_1^d(p)$ checks an increasing number of moment conditions as $T \to \infty$. In fact, a plausible alternative approach that is closer in spirit to the Wooldridge (1990a) test is to consider the following sample moment condition:[10]

$$\hat{\Gamma}(v) = \frac{1}{T}\sum_{t=p+1}^{T} i\hat{\varepsilon}_t\hat{\Lambda}_t(v), \qquad v \in \mathbb{R},$$

where $\hat{\Lambda}_t(v) = \hat{\Psi}_t(v) - \hat{G}_t'\hat{B}(v)$, $\hat{\Psi}_t(v) = [\hat{\psi}_{t-1}(v),\ldots,\hat{\psi}_{t-p}(v)]'$, $\hat{B}(v) = (\sum_{t=p+1}^T \hat{G}_t\hat{G}_t')^{-1}\sum_{t=p+1}^T \hat{G}_t\hat{\Psi}_t(v)$ is the OLS coefficient of regressing $\hat{\Psi}_t(v)$ on $\hat{G}_t$, and p is a fixed lag order. Straightforward algebra shows that

$$\sqrt{T}\hat{\Gamma}(v) = T^{-1/2}\sum_{t=p+1}^{T} i\varepsilon_t\Lambda_t(v) + O_P(T^{-1/2}),$$

where $\Lambda_t(v) = \Psi_t(v) - G_t'B(v)$, where $\Psi_t(v) = [\psi_{t-1}(v),\ldots,\psi_{t-p}(v)]'$ and $B(v) = [E(G_tG_t')]^{-1}E[G_t\Psi_t(v)]$. It follows that $\sqrt{T}\hat{\Gamma}(v) \Rightarrow Z(v)$ on every compact set on $\mathbb{R}$, where $\Rightarrow$ denotes weak convergence, and $Z(v)$ is a complex-valued Gaussian process with mean 0 and variance-covariance kernel $\text{cov}[Z(v_1), Z(v_2)] = E[\varepsilon_t^2 \Lambda_t(v_1)\Lambda_t^*(v_2)']$ for $v_1, v_2 \in \mathbb{R}$. To construct a test statistic, we can define a quadratic form

$$\hat{W} = T\int \hat{\Gamma}(v)'\hat{V}^{-1}(v)\hat{\Gamma}^*(v)\, dW(v),$$

where $\hat{V}(v) = T^{-1}\sum_{t=p+1}^{T} \hat{\varepsilon}_t^2 \hat{\Lambda}_t(v)\hat{\Lambda}_t^*(v)'$. Under suitable regularity conditions and using the continuous mapping theorem, we expect that under $\mathbb{H}_0$,

$$\hat{W} \xrightarrow{d} \int \chi_p^2(v)\, dW(v),$$

where $\chi_p^2(v) \equiv Z(v)'V^{-1}(v)Z^*(v)$ may be called a chi-squared process (see Hansen, 1996, p. 417), with $V(v) = E[\varepsilon_t^2 \Lambda_t(v)\Lambda_t^*(v)']$. This test is also robust to conditional heteroskedasticity of unknown form. Unfortunately, the asymptotic distribution of $\int \chi_p^2(v)\, dW(v)$ is not distribution free; it depends on the data generating process (DGP) and cannot be tabulated. This asymptotic distribution, however, can be consistently approximated using the Hansen (1996) resampling method. Because the asymptotic analysis is rather involved and the simulation study is computationally intensive, we defer the investigation of this approach to future research.

3. ASYMPTOTIC DISTRIBUTION

In Sections 3 and 4, we will compare the asymptotic properties of the modified test $\hat{M}_1^d(p)$ in (2.10) and the unmodified test $\hat{M}_1(p)$ in (2.6) under $\mathbb{H}_0$ and $\mathbb{H}_A$, respectively. To derive the null limiting distribution of $\hat{M}_1^d(p)$, we first give some regularity conditions.

Assumption A.1. $\{Y_t\}$ 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 $\{Y_{t-j}, j > 0\}$ in addition to current and lagged exogenous variables $\{Z_{t-j}, j \geq 0\}$.

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

Assumption A.3. Let $I_t^\dagger$ be an observed information set available at time t that may contain some assumed initial values. Then

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left\{E\left[\sup_{\theta\in\Theta}|g(I_{t-1}^\dagger,\theta)-g(I_{t-1},\theta)|\right]^4\right\}^{1/4}\le C,\quad \text{and}$$

$$\lim_{T\to\infty}\sum_{t=1}^{T}\left\{E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}g(I_{t-1}^\dagger,\theta)-\frac{\partial}{\partial\theta}g(I_{t-1},\theta)\right\|\right]^2\right\}^{1/2}\le C.$$

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

Assumption A.5. Put $\varepsilon_t\equiv Y_t-g(I_{t-1},\theta_0)$ and $G_t\equiv(\partial/\partial\theta)g(I_{t-1},\theta_0)$. Then $\{\varepsilon_t,G_t'\}'$ is a strictly stationary α-mixing process with α-mixing coefficient $\alpha(j)$ satisfying $\sum_{j=-\infty}^{\infty}j^2\alpha(j)^{(\nu-1)/\nu}<\infty$ for some $\upsilon>1$. In addition, $E(\varepsilon_t^2)=\sigma^2$ and $E(\varepsilon_t^4)\le C$.

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

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

Assumption A.8. For each sufficiently large integer q, there exists a strictly stationary process $\{\varepsilon_{q,t}\}$ such that as $q\to\infty$, $\varepsilon_{q,t}$ is independent of I_{t-q-1} for each t, $E(\varepsilon_{q,t}|I_{t-1})=0$ *a.s.*, $E(\varepsilon_t-\varepsilon_{q,t})^4\le Cq^{-2\kappa}$ for some constant $\kappa\ge 1$, and $E(\varepsilon_{q,t}^4)\le C$.

Assumption A.1 imposes a strict stationarity condition on the process $\{Y_t\}$. The existence of the conditional mean μ_t can be ensured by assuming that $E(Y_t^2)<\infty$. Assumption A.2 is a standard regularity condition on the conditional mean model $g(I_{t-1},\theta)$. For a (static or dynamic) linear regression model $g(I_{t-1},\theta)=X_t'\theta$, where $X_t\in I_{t-1}$ is finite-dimensional, it suffices if $E\|X_t\|^4\le C$ and $E(X_tX_t')$ is nonsingular. Assumption A.2 covers many stationary nonlinear time series conditional mean models, such as nonlinear moving-average, bilinear, exponential, Markov regime-switching, smooth transition, and Poisson jump autoregressive models. It also covers threshold autoregressive models with known thresholds. An example is the class of self-exciting autoregressive threshold models for the U.S. economy, where the recession and the expansion are defined as the gross domestic product (GDP) growth rate being larger or smaller than zero (e.g., Potter, 1995). However, Assumption A.2 rules out the autoregressive threshold models with unknown thresholds as 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 regularity conditions, but we do not attempt to justify this here, which is beyond the scope of this paper. We note that Assumption A.2(c) was not needed for

$\hat{M}_1(p)$, but it is needed for $\hat{M}_1^d(p)$ because we use a sequence of auxiliary OLS regressions.

Assumption A.3 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 completely observable. Because of the truncation, one may have to assume some initial values in estimating the model $g(I_{t-1},\theta)$. Assumption A.3 ensures that the use of initial values, if any, has no impact on the limiting distribution of $\hat{M}_1^d(p)$. For instance, consider an ARMA(1,1) model:

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

where $|\alpha| \le \bar{\alpha} < 1$ and $|\beta| \le \bar{\beta} < \infty$. Here $I_{t-1} = \{Y_{t-1}, Y_{t-2}, \ldots\}$ but $I_{t-1}^{\dagger} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_1, \bar{\varepsilon}_0\}$, and $\bar{\varepsilon}_0$ is an initial value assumed for ε_0. By recursive substitution, Hong and Lee (2005) showed

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

Similarly we can show that the information truncation condition for $\{(\partial/\partial\theta)g(I_{t-1}^{\dagger},\theta) - (\partial/\partial\theta)g(I_{t-1},\theta)\}$ in Assumption A.3 holds for the ARMA(1,1) model. This condition was not needed for $\hat{M}_1(p)$ but is needed for $\hat{M}_1^d(p)$.

Assumption A.4 requires a $\sqrt{T}$-consistent estimator $\hat{\theta}$, which need not be asymptotically most efficient. It can be a conditional least squares estimator or a conditional quasi–maximum likelihood estimator. Also, we need not know the asymptotic expansion structure of $\hat{\theta}$, because the sampling variation in $\hat{\theta}$ does not affect the asymptotic distribution of $\hat{M}_1^d(p)$. These features are similar in spirit to the Wooldridge (1990a) modified *m*-tests. Assumption A.5 imposes mixing conditions on $\{\varepsilon_t, G_t'\}'$, which restrict the degree of the serial dependence in $\{\varepsilon_t, G_t'\}'$. The mixing condition is suitable and convenient for nonlinear time series analysis. For more discussion on mixing conditions, see (e.g.) White (2001).

Assumption A.6 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 $\hat{S}^{(0,1,0)}(\omega,0,v)$ in (2.9) vanishes to 0 as $T \to \infty$. The tail condition on $k(\cdot)$ requires that $k(z)$ decays to zero sufficiently fast as $|z| \to \infty$. It is more stringent than that imposed in Hong and Lee (2005). It implies that $\int_0^\infty (1+z^2)|k(z)|\,dz < \infty$. This condition rules out the Daniell and quadratic spectral kernels, whose $b = 2$.[11] However, it includes all kernels with bounded support, such as the Bartlett and Parzen kernels, because they have $b = \infty$. Assumption A.7 is a condition on the weighting function $W(\cdot)$ for the transform parameter v. It is satisfied by the c.d.f. of any symmetric continuous distribution with a finite fourth moment.

Assumption A.8 is required only under $\mathbb{H}_0$. It assumes that when q is sufficiently large, the m.d.s. $\{\varepsilon_t\}$ can be approximated by a q-dependent m.d.s. process $\{\varepsilon_{q,t}\}$ arbitrarily well. Horowitz (2003) imposed a similar condition in a different context. Because $\{\varepsilon_t\}$ is an m.d.s. under $\mathbb{H}_0$, Assumption A.8 essentially imposes restrictions on the serial dependence in the higher order moments of $\{\varepsilon_t\}$. It holds trivially when $\{\varepsilon_t\}$ is a q_0-dependent process with an arbitrarily large but fixed order q_0. It also covers many non-Markovian processes. For example, Hong and Lee (2005) showed that Assumption A.8 holds for a threshold GARCH(1,1) error process that includes a standard GARCH process as a special case:

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

provided $\rho < 1$, where $\rho \equiv \alpha + \beta^+ + (\beta^- - \beta^+) E[z_t^2 \mathbf{1}(z_t \le 0)]$ and $\mathbf{1}(\cdot)$ is an indicator function. It also holds for a general stochastic volatility process:

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

where $\sum_{j=1}^{\infty} \alpha_j^2 < \infty$, $E(z_t^4) < \infty$, and $\{z_t\}$ and $\{\eta_t\}$ may not be independent of each other.

We now state the asymptotic distribution of the $\hat{M}_1^d(p)$ test under $\mathbb{H}_0$.

THEOREM 1. *Suppose Assumptions A.1–A.8 hold and* $p = cT^{\lambda}$ *for* $0 < \lambda < (3 + (1/(4b - 2)))^{-1}$ *and* $0 < c < \infty$. *Then* $\hat{M}_1^d(p) - \hat{M}_1(p) \xrightarrow{p} 0$ *and* $\hat{M}_1^d(p) \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0$.

As with the original test $\hat{M}_1(p)$, we obtain the convenient asymptotic $N(0,1)$ distribution for the modified test $\hat{M}_1^d(p)$, but the latter is expected to have a better finite-sample performance. Theorem 1 also implies that $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ are asymptotically equivalent under $\mathbb{H}_0$. The asymptotic equivalence holds even if the orthogonality condition that $T^{-1/2} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t \xrightarrow{p} 0$ fails. This orthogonality condition is needed for the Wooldridge (1990a) modified and unmodified parametric m-tests to be asymptotically equivalent under the null hypothesis (see Wooldridge, 1990a, Lem. 2.2, pp. 28–29). It will hold when one uses the nonlinear least squares estimator

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{t=1}^{T} [Y_t - g(I_{t-1}^{\dagger}, \theta)]^2.$$

The asymptotic equivalence between $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ under $\mathbb{H}_0$ has an important implication. Although we do not formally analyze it here, we expect that the asymptotic equivalence between $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$ will continue to hold under a suitable class of local alternatives to $\mathbb{H}_0$ (for a similar discussion on unmodified and modified parametric m-tests, see Wooldridge, 1990a, p. 29). In other words, $\hat{M}_1^d(p)$ will be asymptotically as powerful as $\hat{M}_1(p)$ under a class of local alternatives.

We now summarize the procedures to implement the modified test $\hat{M}_1^d(p)$ as follows.

Step 1. Obtain a $\sqrt{T}$-consistent estimator $\hat{\theta}$ (e.g., the nonlinear least squares estimator) for the conditional mean model $g(I_{t-1}, \theta)$ and save the estimated residual $\hat{\varepsilon}_t$.

Step 2. Compute $\hat{G}_t = (\partial/\partial\theta) g(I_{t-1}^{\dagger}, \hat{\theta})$. For a linear regression model $g(I_{t-1}, \theta) = X_t' \theta$, we have $\hat{G}_t = X_t$, the regressor vector.

Step 3. For each lag order j from 1 to $T-1$, run an OLS regression of $\hat{\psi}_{t-j}(v)$ on $\hat{G}_t$, where we set $\hat{\psi}_t(v) = 0$ for $t \leq 0$.[12] Save the estimated residual $\hat{h}_{t-j}(v)$. If the kernel $k(\cdot)$ has a bounded support (i.e., $k(z) = 0$ if $|z| > 1$), then it suffices to run regressions for j from 1 to p.

Step 4. Compute the modified test statistic $\hat{M}_1^d(p)$ in (2.10).

Step 5. Compare $\hat{M}_1^d(p)$ with an upper tailed $N(0,1)$ critical value (e.g., 1.645 at the 5% level) and reject $\mathbb{H}_0$ at a suitable significance level if $\hat{M}_1^d(p)$ is larger than the critical value.

We note that for a static conditional mean model $g(I_{t-1}, \theta) = g(X_t, \theta)$, where X_t is a strictly exogenous random vector independent of innovations $\{\varepsilon_t\}$, we need not use the Wooldridge (1990a) device because, as pointed out earlier, the impact of parameter estimation uncertainty has been a rather small order in this case and the Wooldridge (1990a) device cannot further reduce the order of magnitude for the higher order terms in $\hat{M}_1(p)$ that are associated with $\hat{\theta}$. However, for a dynamic conditional mean model $g(I_{t-1}, \theta) = g(X_t, \theta)$, where X_t contains lagged dependent variables and/or lagged innovations, the Wooldridge (1990a) device can reduce the order of magnitude of the higher order terms in $\hat{M}_1^d(p)$ that are associated with $\hat{\theta}$, thus achieving a better normal approximation in small and finite samples.

In this paper we have focused on time series conditional mean models with additive errors in (2.1). In fact, our approach is also applicable to a time series conditional mean model with multiplicative errors:

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

where $\{Y_t\}$ is a nonnegative stochastic time series, $g(I_{t-1}, \theta)$ is a parametric model for the conditional mean $E(Y_t | I_{t-1})$, and $\varepsilon_t(\theta)$ is a nonnegative multiplicative error. Examples of the model in (3.1) include the Engle and Russell (1998) autoregressive conditional duration model for duration times of financial events and its various extensions. When and only when the model $g(I_{t-1},\theta)$ is correctly specified for $E(Y_t|I_{t-1})$, we have

$$E\{[\varepsilon_t(\theta_0)-1]|I_{t-1}\}=0 \ a.s. \quad \text{for some } \theta_0\in\Theta.$$

The generalized spectral derivative tests can be applied to the model in (3.1), with the estimated residual $\hat{\varepsilon}_t$ and the gradient $\hat{G}_t$ used in $\hat{M}_1^d(p)$ being replaced by $\hat{\varepsilon}_t = Y_t/g(I_{t-1}^{\dagger},\hat{\theta})-1$ and $\hat{G}_t=(\partial/\partial\theta)\ln g(I_{t-1}^{\dagger},\hat{\theta})$.

4. ASYMPTOTIC POWER

To investigate the power property of $\hat{M}_1^d(p)$, particularly the impact of the auxiliary regressions on the asymptotic power of $\hat{M}_1^d(p)$, we consider the asymptotic behavior of $\hat{M}_1^d(p)$ under $\mathbb{H}_A$. For this purpose, we define the population modified generalized spectrum

$$S(\omega,u,v)=\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j(u,v)e^{-ij\omega},\qquad \omega\in[-\pi,\pi],\qquad u,v\in\mathbb{R}, \tag{4.1}$$

where $\gamma_j(u,v)=\text{cov}[e^{iu\varepsilon_t},h_{t-j}(v)]$, $h_{t-j}(v)=\psi_{t-j}(v)-G_t'\beta_j(v)$, $G_t=(\partial/\partial\theta)g(I_{t-1},\theta_0)$,

$$\beta_j(v)=[E(G_tG_t')]^{-1}E[G_t\psi_{t-j}(v)], \tag{4.2}$$

and $\psi_t(v)=e^{iv\varepsilon_t}-\varphi(v)$. We also define the population modified generalized "flat" spectrum:

$$S_0(\omega,u,v)=\frac{1}{2\pi}\gamma_0(u,v),\qquad \omega\in[-\pi,\pi],\qquad u,v\in\mathbb{R}.$$

Then the partial derivatives of the modified generalized spectrum are $S^{(0,1,0)}(\omega,0,v)\equiv(\partial/\partial u)S(\omega,u,v)|_{u=0}$ and $S_0^{(0,1,0)}(\omega,0,v)\equiv(\partial/\partial u)S_0(\omega,u,v)|_{u=0}$, provided these partial derivatives exist. With these notations, we can state Theorem 2.

THEOREM 2. *Suppose Assumptions A.1–A.7 hold and* $p=cT^{\lambda}$ *for* $0<\lambda<\frac{1}{2}$ *and* $0<c<\infty$. *Then as* $T\to\infty$,

$$\begin{aligned}(p^{1/2}/T)\hat{M}_1^d(p) &\xrightarrow{p}\left[2D_d\int_0^\infty k^4(z)\,dz\right]^{-1/2}\\ &\quad\times\pi\iint_{-\pi}^{\pi}|S^{(0,1,0)}(\omega,0,v)-S_0^{(0,1,0)}(\omega,0,v)|^2\,d\omega dW(v)\\ &=\left[2D_d\int_0^\infty k^4(z)\,dz\right]^{-1/2}\sum_{j=1}^{\infty}\int|\gamma_j^{(1,0)}(0,v)|^2\,dW(v),\end{aligned}$$

where $D_d \equiv \sigma^4 \sum_{j=-\infty}^{\infty} \int\int |\rho_j(u,v)|^2 \, dW(u)\, dW(v)$ *and* $\rho_j(u,v) = \mathrm{cov}[h_t(u), h_{t-j}(v)]$.

The constant D_d takes into account the impact of serial dependence in conditioning functions $\{h_{t-j}(v), j > 0\}$, which generally exists even under $\mathbb{H}_0$, due to the presence of serial dependence in the conditional variance and higher order moments of $\{\varepsilon_t\}$. It has also taken into account the impact of the auxiliary regressions on the behavior of $\hat{M}_1^d(p)$. For comparison, we have from Hong and Lee (2005, Thm. 2) that for the unmodified test,

$$(p^{1/2}/T)\hat{M}_1(p) \xrightarrow{p} \left[2D\int_0^{\infty} k^4(z)\,dz\right]^{-1/2} \times \pi \iint_{-\pi}^{\pi} |f^{(0,1,0)}(\omega,0,v) - f_0^{(0,1,0)}(\omega,0,v)|^2 \, d\omega dW(v)$$
$$= \left[2D\int_0^{\infty} k^4(z)\,dz\right]^{-1/2} \sum_{j=1}^{\infty} \int |\sigma_j^{(1,0)}(0,v)|^2 \, dW(v),$$

where $D \equiv \sigma^4 \sum_{j=-\infty}^{\infty} \int\int |\sigma_j(u,v)|^2 \, dW(u)\, dW(v)$. Clearly, $\hat{M}_1^d(p)$ and $\hat{M}_1(p)$ are not asymptotically equivalent under $\mathbb{H}_A$ because they do not converge to the same probability limit after being multiplied by the rate $p^{1/2}/T$. Unlike the case under $\mathbb{H}_0$, where the auxiliary regressions have no impact on the asymptotic distribution of $\hat{M}_1^d(p)$, the auxiliary regressions have impact on the probability limit of $(p^{1/2}/T)\hat{M}_1^d(p)$ under $\mathbb{H}_A$.

We now discuss how the auxiliary regressions may affect the power of $\hat{M}_1(p)$. 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 $\mathbb{R}$. It follows that $\mathrm{P}[\hat{M}_1(p) > C(T)] \to 1$ for any sequence of constants $C(T) = o(T/p^{1/2})$, and so the unmodified test $\hat{M}_1(p)$ has asymptotic unit power at any given significance level $\alpha \in (0,1)$, whenever $E(\varepsilon_t|\varepsilon_{t-j})$ is nonzero at some lag $j > 0$. This is the reason why $\hat{M}_1(p)$ has omnibus power against a wide variety of linear and nonlinear alternatives with unknown lag structure, as is confirmed in the Hong and Lee (2005) simulation. It avoids the blindness of searching for different alternatives when one has no prior information.

Theorem 2 indicates that the power of $\hat{M}_1^d(p)$ depends on whether $\gamma_j^{(1,0)}(0,v) \neq 0$ at least for some $j > 0$ under $\mathbb{H}_A$. Note that $\gamma_j^{(1,0)}(0,v) \neq \sigma_j^{(1,0)}(0,v)$ generally under $\mathbb{H}_A$. However, if we have either (i) $E(G_t\varepsilon_t) = 0$ or (ii) $\beta_j(v) = 0$ for all $j > 0$ under $\mathbb{H}_A$, then $\gamma_j^{(1,0)}(0,v) = \sigma_j^{(1,0)}(0,v)$ for all $v \in \mathbb{R}$. In these cases, $\hat{M}_1^d(p)$ has the same consistency property as $\hat{M}_1(p)$, although their probability limits still may be different, because of the fact that the denominator D_d depends on $\beta_j(v)$ when $\beta_j(v) \neq 0$ at least for some $j > 0$. The case that $\beta_j(v) = 0$ for all $j > 0$ can arise when $g(I_{t-1},\theta_0) = g(X_t,\theta_0)$, where X_t is a strictly exogenous vector independent of innovations $\{\varepsilon_t\}$. The case that $E(G_t\varepsilon_t) = 0$ can arise under $\mathbb{H}_A$ even when $g(I_{t-1},\theta_0)$ contains lagged dependent variables and/or lagged innovations. In particular, when $\hat{\theta}$ is a nonlinear least squares estimator, i.e., $\hat{\theta} = \arg\min_{\theta\in\Theta}\sum_{t=1}^{T}[Y_t - g(I_{t-1},\theta)]^2$, then $\theta_0 = p\lim\hat{\theta}$ will satisfy the condition that $E(G_t\varepsilon_t) = E[(\partial/\partial\theta)g(I_{t-1},\theta_0)\varepsilon_t] = 0$ under $\mathbb{H}_A$.

When $E(G_t\varepsilon_t) \neq 0$ and $\beta_j(v) \neq 0$ at least for some $j > 0$, we generally have $\gamma_j^{(1,0)}(0,v) \neq 0$ if $\sigma_j^{(1,0)}(0,v) \neq 0$. However, there exists a certain model misspecification against which the modified test $\hat{M}_1^d(p)$ has no power. This arises when $\gamma_j^{(1,0)}(0,v) = 0$ for all $j > 0$ but $\sigma_j^{(1,0)}(0,v) \neq 0$ for some $j > 0$. Let

$$\alpha \equiv [E(G_tG_t')]^{-1}E(G_t\varepsilon_t)$$

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

$$\sigma_j^{(1,0)}(0,v) = i\alpha' E(G_tG_t')\beta_j(v) = i\operatorname{cov}[\alpha' G_t, G_t'\beta_j(v)] \quad \text{for all } j > 0,$$

i.e., if and only if the covariance between ε_t and $e^{iv\varepsilon_{t-j}}$ coincides with the covariance between their linear projections onto G_t. This occurs when the neglected dynamics in mean takes the form of $E(\varepsilon_t|I_{t-1}) = \alpha'[G_t - E(G_t)]$. In this (pathological) case, the modified test $\hat{M}_1^d(p)$ has no power. This is the price that we have to pay when using the Wooldridge (1990a) device, as is also the case in parametric testing (for discussion, see White, 1994, Ch. 9). However, we emphasize that the gain in the size improvement from using Wooldridge's device for our tests overwhelms the possible power loss in detecting misspecification in the direction of the gradient G_t. More importantly, if the nonlinear least squares estimator is used, $\hat{M}_1^d(p)$ will be able to detect such pathological misspecification and achieve the same consistency property as the original test $\hat{M}_1(p)$.

Because existing tests for time series conditional mean models only consider a fixed order lag, they can easily miss misspecifications at the 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, because of the loss of a large number of degrees of freedom. This power loss is greatly alleviated for our tests as a result of the use of $k^2(\cdot)$. Most nonuniform kernels discount higher order lags. This enhances good power against the alternatives whose serial dependence in mean 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 shared by popular chi-square-type tests with a large number of lags, which essentially give equal weighting to each lag. Equal weighting is not fully efficient when a large number of lags is used.

Once the model $g(I_{t-1},\theta)$ is rejected by $\hat{M}_1^d(p)$, one may want to go further to explore possible sources of model misspecification in mean. For this purpose, we can further differentiate the modified generalized spectral derivative $S^{(0,1,0)}(\omega,0,v)$ with respect to v at 0 and construct a sequence of tests similar in spirit to $\hat{M}_1^d(p)$. Specifically, the partial derivatives

$$\gamma_j^{(1,l)}(0,0) \equiv \frac{\partial^l}{\partial v^l}\gamma_j^{(1,0)}(0,v)|_{v=0}$$
$$= \text{cov}\left[i\varepsilon_t,(i\varepsilon_{t-j})^l - G_t'\frac{d^l}{dv^l}\beta_j(0)\right], \qquad l=0,1,2,\ldots,$$

can be used. For $l = 1,2,3,4$, tests based on these derivatives can check whether there exist linear correlation, ARCH-in-mean, skewness-in-mean, and kurtosis-in-mean effects, respectively. ARCH-in-mean effects are important in finance, and the recent literature has also documented time-varying skewness and kurtosis and their economic significance in asset pricing (e.g., Harvey and Siddique, 1999, 2000).

5. MONTE CARLO EVIDENCE

The results of Theorems 1 and 2 are asymptotic. Nothing is known about the finite-sample performance of the modified test $\hat{M}_1^d(p)$ relative to the unmodified test $\hat{M}_1^d(p)$. We now investigate their finite-sample performance.

5.1. Simulation Design

5.1.1. Size. To examine the sizes of the tests under $\mathbb{H}_0$, we consider the following AR(d) processes:

$$Y_t = \sum_{j=1}^{d} 0.5^j Y_{t-j} + \varepsilon_t,$$

where (i) $\varepsilon_t = z_t$ or (ii) $\varepsilon_t = h_t^{1/2} z_t$, $h_t = 0.43 + 0.57\varepsilon_{t-1}^2$, where $\{z_t\} \sim i.i.d.N(0,1)$.[13] Under (i), $\{\varepsilon_t\} \sim i.i.d.$, whereas under (ii), $\{\varepsilon_t\}$ is an ARCH(1) process.

The null conditional mean model for Y_t is an AR(d) model with intercept:

$$g(I_{t-1},\theta) = \theta_1 + \sum_{j=1}^{d}\theta_{j+1}Y_{t-j}. \qquad \textbf{(5.1)}$$

To examine the impact of increasing the number of estimated autoregressive parameters, we consider $d = 1,2,3,4$, respectively. The OLS estimator $\hat{\theta}$ is consistent for parameter $\theta_0 \equiv (\theta_1,\ldots,\theta_{d+1})'$. The model error $\{\varepsilon_t(\theta_0)\}$ is conditionally homoskedastic under the i.i.d. innovations and is conditionally heteroskedastic under the ARCH innovations. This allows us to examine the robustness of the tests to conditional heteroskedasticity. We have chosen ARCH parameter values such that $E(\varepsilon_t^4) < \infty$, thus satisfying Assumption A.5.[14] To examine the size, we consider three sample sizes: $T = 100$, 250, and 500. For each T, we generate 1,000 data sets using a GAUSS Windows version 5.0 random number generator on a personal computer. For each iteration, we first generate $T + 100$ observations and then discard the first 100 to reduce the impact of some initial values.

5.1.2. Power. Next, we examine the power of the tests for neglected nonlinearity or dynamic misspecification, i.e., lag order misspecification in mean. Following Hong and Lee (2005), we consider the following data 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)]: $Y_t = \begin{cases} 0.5Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} \le 0, \\ -0.5Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} > 0, \end{cases}$

DGP P.5 [STAR(1)]: $Y_t = 1 - 0.5Y_{t-1} - (4 + 0.4Y_{t-1})G(2Y_{t-1}) + \varepsilon_t,$ 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 $i.i.d.N(0,1)$. These DGPs are discussed in Hong and Lee (2005). We will examine the power of the modified test $\hat{M}_1^d(p)$ relative to the unmodified test $\hat{M}_1(p)$ for two sample sizes: $T = 100$ and 250. For each T, we generate 500 data sets.

To compute $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$, we use the $N(1,0)$ c.d.f. truncated on $[-3,3]$ for the weighting function $W(\cdot)$, and we use the Parzen kernel

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

which has a bounded support and is computationally efficient. For the choice of lag order p, we use a data-driven lag order $\hat{p}_0$ via the plug-in method described in Hong and Lee (2005, Sect. 6), with the Bartlett kernel $\bar{k}(z) = (1 - |z|)\mathbf{1}(|z| \le 1)$ used in the preliminary generalized spectral derivative estimators. To certain extent, the data-driven lag order $\hat{p}_0$ lets data tell an appropriate lag order, but it still involves the choice of the preliminary bandwidth $\bar{p}$, which is somewhat arbitrary. To examine the impact of the choice of the preliminary bandwidth $\bar{p}$, we consider $\bar{p} = 10, 15, 20, 25, 30$, which cover a sufficiently wide range of preliminary lag orders.

5.2. Monte Carlo Evidence

Tables 1 and 2 report the empirical rejection rates of the tests under $\mathbb{H}_0$ at the 10% and 5% significance levels, using the asymptotic theory. We first consider the unmodified test $\hat{M}_1(\hat{p}_0)$. When the DGP is an AR(d) process with i.i.d. errors, $\hat{M}_1(\hat{p}_0)$ underrejects $\mathbb{H}_0$ severely at both the 10% and 5% levels (particularly for larger d), even when $T = 500$. When the DGP is an AR(d) process with ARCH errors, $\hat{M}_1(\hat{p}_0)$ also shows underrejection (particularly when $T = 100$), but the sizes are better than under DGPs with i.i.d. errors when $T \geq 250$, and they improve as T increases. Overall, as the lag order d of the AR model increases, $\hat{M}_1(\hat{p}_0)$ displays more severe underrejection (except for the case at the 5% level and when $T = 500$), thus confirming our conjecture.

To investigate whether the underrejection of $\hat{M}_1(\hat{p}_0)$ is due to the impact of parameter estimation uncertainty, we also report the rejection rates of the infeasible test statistic $\widetilde{M}_1(\hat{p}_0)$ of (2.13) that uses the true errors $\{\varepsilon_t\}$ rather than the estimated model residuals $\{\hat{\varepsilon}_t\}$. When $\{\varepsilon_t\}$ is i.i.d., $\widetilde{M}_1(\hat{p}_0)$ outperforms $\hat{M}_1(\hat{p}_0)$ for all sample sizes T, and its sizes are reasonable and are robust to the choice of the preliminary lag order $\bar{p}$. Note that the rejection rates of $\widetilde{M}_1(\hat{p}_0)$ remain unchanged when the order of the AR(d) model increases, because $\widetilde{M}_1(\hat{p}_0)$ does not use the estimated model residuals $\{\hat{\varepsilon}_t\}$. When $\{\varepsilon_t\}$ is ARCH(1), $\widetilde{M}_1(\hat{p}_0)$ shows some underrejections when $T = 100$, but its sizes become reasonable when $T \geq 250$. For all sample sizes T, $\widetilde{M}_1(\hat{p}_0)$ outperforms $\hat{M}_1(\hat{p}_0)$. These results confirm our theory that parameter estimation uncertainty has nontrivial impact on $\hat{M}_1(\hat{p}_0)$ in finite samples. In particular, parameter estimation is like a calibration that makes the estimated model residuals look more like an m.d.s., leading to underrejection of the test.

For each sample size T and each preliminary lag order $\bar{p}$, the means of the data-driven lag $\hat{p}_0$ are the same or very similar for all $\hat{M}_1(\hat{p}_0), \widetilde{M}_1(\hat{p}_0)$, and $\hat{M}_1^d(\hat{p}_0)$, regardless of the order of the AR(d) model and the types (i.i.d. or ARCH(1)) of the errors. When the preliminary lag $\bar{p}$ changes from 10 to 30, the means of $\hat{p}_0$ range from 4.6 to 7.2 for $T = 100$, from 5.5 to 7.2 for $T = 250$, and from 6.2 to 7.3 for $T = 500$. These results indicate relative robustness of $\hat{p}_0$ to the choice of $\bar{p}$.[15]

We now turn to the modified test $\hat{M}_1^d(\hat{p}_0)$ to examine whether the Wooldridge (1990a) device can effectively reduce the impact of parameter estimation uncertainty on the sizes of the test. For $T = 100$, $\hat{M}_1^d(\hat{p}_0)$ shows some overrejections when the order d of the AR(d) model is relatively large. However, it has reasonable sizes under all DGPs for $T \geq 250$. The $\hat{M}_1^d(\hat{p}_0)$ test performs

TABLE 1. Empirical size of tests under i.i.d. errors

T	$T=100$						$T=250$						$T=500$					
	10%			5%			10%			5%			10%			5%		
$\bar{p}$	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d
AR(1): $Y_t = 0.5Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0,1)$																		
10	3.0	8.3	9.4	1.5	5.4	5.9	4.4	9.8	10.5	2.6	6.7	6.1	4.6	11.2	11.0	3.8	6.0	7.2
15	3.3	7.9	9.3	1.8	5.5	6.0	4.4	9.8	10.5	2.6	6.7	6.1	4.6	11.2	11.0	3.8	6.0	7.2
20	3.6	8.0	8.9	1.8	5.3	5.2	4.5	10.2	10.1	2.6	6.7	5.7	4.6	11.2	11.0	3.8	6.0	7.2
25	3.6	7.9	9.0	1.8	5.3	4.8	4.1	9.8	9.4	2.8	6.2	5.5	4.7	11.2	11.0	4.0	6.0	7.2
30	3.9	7.8	8.9	2.0	5.4	4.8	4.2	10.5	8.5	2.6	5.7	5.6	4.5	10.6	10.7	4.0	6.4	6.9
AR(2): $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0,1)$																		
10	1.3	8.3	11.3	0.5	5.4	7.2	1.6	9.8	9.1	0.9	6.7	6.0	2.2	11.2	11.5	1.1	6.0	7.8
15	1.2	7.9	11.0	0.5	5.5	7.3	1.6	9.8	9.1	0.9	6.7	6.0	2.2	11.2	11.5	1.1	6.0	7.8
20	1.2	8.0	11.4	0.6	5.3	6.5	1.8	10.2	9.2	0.9	6.7	5.9	2.2	11.2	11.5	1.1	6.0	7.8
25	1.4	7.9	11.0	0.6	5.3	6.0	1.8	9.8	8.6	0.9	6.2	6.0	2.4	11.2	11.4	1.1	6.0	7.6
30	1.7	7.8	10.2	0.6	5.4	5.5	1.9	10.5	8.6	1.0	5.7	5.6	2.7	10.6	11.0	1.3	6.4	7.4
AR(3): $Y_t = \sum_{j=1}^{3} 0.5^j Y_{t-j} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0,1)$																		
10	0.9	8.3	10.8	0.2	5.4	7.3	0.8	9.8	8.7	0.5	6.7	6.1	0.9	11.2	9.1	0.3	6.0	6.2
15	0.8	7.9	10.7	0.2	5.5	7.5	0.8	9.8	8.7	0.5	6.7	6.1	0.9	11.2	9.1	0.3	6.0	6.2
20	0.8	8.0	11.0	0.2	5.3	7.2	0.8	10.2	8.6	0.5	6.7	5.9	0.9	11.2	9.1	0.3	6.0	6.2
25	0.7	7.9	10.9	0.1	5.3	7.2	0.8	9.8	8.1	0.6	6.2	5.7	1.0	11.2	9.1	0.4	6.0	6.3
30	0.7	7.8	10.8	0.1	5.4	7.0	1.0	10.5	8.0	0.5	5.7	5.3	1.1	10.6	9.6	0.6	6.4	6.4
AR(4): $Y_t = \sum_{j=1}^{4} 0.5^j Y_{t-j} + \varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0,1)$																		
10	0.7	8.3	12.3	0.3	5.4	9.5	0.9	9.8	9.6	0.1	6.7	6.5	0.5	11.2	11.6	0.1	6.0	7.5
15	0.7	7.9	12.6	0.3	5.5	9.7	0.9	9.8	9.6	0.1	6.7	6.5	0.5	11.2	11.6	0.1	6.0	7.5
20	0.7	8.0	12.6	0.2	5.3	9.4	0.8	10.2	9.9	0.1	6.7	6.4	0.5	11.2	11.7	0.1	6.0	7.5
25	0.6	7.9	12.3	0.2	5.3	9.2	0.7	9.8	10.1	0.1	6.2	5.9	0.6	11.2	11.5	0.1	6.0	7.3
30	0.5	7.8	12.3	0.2	5.4	8.8	0.6	10.5	9.8	0.1	5.7	5.5	0.5	10.6	11.4	0.1	6.4	7.2

Notes: (i) 1,000 iterations; (ii) $\hat{M}_1(\hat{p}_0)$, $\hat{M}_1^d(\hat{p}_0)$, the original and modified generalized spectral tests derived under time-varying higher moments, respectively, $\widetilde{M}_1(\hat{p}_0)$, the infeasible original generalized spectral test; (iii) $\bar{p}$, the preliminary lag order used in a plug-in method to choose a data-dependent lag order $\hat{p}_0$. The Parzen kernel is used for $\hat{M}_1(\hat{p}_0)$, $\hat{M}_1^d(\hat{p}_0)$, and $\widetilde{M}_1(\hat{p}_0)$.

Table 2. Empirical size of tests under ARCH(1) errors

T	$T = 100$						$T = 250$						$T = 500$					
	10%			5%			10%			5%			10%			5%		
$\bar{p}$	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d	M_1	$\widetilde{M}_1$	M_1^d
AR(1): $Y_t = 0.5Y_{t-1} + z_t\sqrt{h_t}$, $h_t = 0.43 + 0.57\varepsilon_{t-1}^2$, $z_t \sim i.i.d.N(0,1)$																		
10	2.5	6.4	8.7	1.5	3.9	5.2	6.2	8.8	9.6	3.4	5.0	6.9	8.2	9.3	9.4	4.2	6.2	5.6
15	2.3	6.3	8.7	1.8	3.7	4.9	6.2	8.8	9.6	3.4	5.0	6.9	8.2	9.3	9.4	4.2	6.2	5.6
20	2.2	6.2	8.4	1.8	3.4	4.4	6.2	8.8	9.4	3.5	4.9	6.8	8.2	9.3	9.4	4.2	6.2	5.6
25	2.0	6.1	7.8	1.8	3.2	3.9	5.6	8.4	9.2	3.2	4.4	6.2	8.3	9.4	9.5	4.3	5.7	5.5
30	1.9	6.0	7.7	2.0	3.1	4.0	5.5	7.7	8.9	3.0	4.3	5.6	8.1	8.9	9.4	4.2	5.6	5.4
AR(2): $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + z_t\sqrt{h_t}$, $h_t = 0.43 + 0.57\varepsilon_{t-1}^2$, $z_t \sim i.i.d.N(0,1)$																		
10	2.5	6.4	8.7	1.5	3.9	7.0	5.4	8.8	9.6	3.5	5.0	6.7	7.5	9.3	9.4	3.6	6.2	6.0
15	2.3	6.3	8.7	1.4	3.7	7.0	5.4	8.8	9.6	3.5	5.0	6.7	7.5	9.3	9.4	3.6	6.2	6.0
20	2.2	6.2	8.4	1.4	3.4	6.5	5.3	8.8	9.4	3.5	4.9	6.6	7.5	9.3	9.4	3.6	6.2	6.0
25	2.0	6.1	7.8	1.3	3.2	6.2	4.9	8.4	9.2	2.9	4.4	6.4	7.5	9.4	9.5	3.7	5.7	6.1
30	1.9	6.0	7.7	1.1	3.1	7.5	4.5	7.7	8.9	2.6	4.3	6.1	7.5	8.9	9.4	4.1	5.6	6.0
AR(3): $Y_t = \sum_{j=1}^{3} 0.5^j Y_{t-j} + z_t\sqrt{h_t}$, $h_t = 0.43 + 0.57\varepsilon_{t-1}^2$, $z_t \sim i.i.d.N(0,1)$																		
10	2.7	6.4	12.9	1.4	3.9	8.3	5.8	8.8	10.8	3.3	5.0	7.2	7.0	9.3	9.3	4.1	6.2	6.8
15	2.5	6.3	12.5	1.4	3.7	8.2	5.8	8.8	10.8	3.3	5.0	7.2	7.0	9.3	9.3	4.1	6.2	6.8
20	2.4	6.2	12.0	1.2	3.4	7.5	5.9	8.8	10.4	3.3	4.9	7.2	7.1	9.3	9.3	4.1	6.2	6.8
25	1.9	6.1	12.1	1.2	3.2	7.6	5.4	8.4	10.4	2.9	4.4	7.0	7.1	9.4	9.5	4.0	5.7	6.5
30	1.8	6.0	11.8	1.0	3.1	7.1	5.4	7.7	10.4	2.7	4.3	6.7	7.1	8.9	9.5	3.9	5.6	6.0
AR(4): $Y_t = \sum_{j=1}^{4} 0.5^j Y_{t-j} + z_t\sqrt{h_t}$, $h_t = 0.43 + 0.57\varepsilon_{t-1}^2$, $z_t \sim i.i.d.N(0,1)$																		
10	2.0	6.4	13.7	1.0	3.9	8.8	4.7	8.8	11.2	2.9	5.0	6.7	6.8	9.3	11.0	5.0	6.2	6.9
15	2.0	6.3	13.6	1.0	3.7	8.7	4.7	8.8	11.2	2.9	5.0	6.7	6.8	9.3	11.0	5.0	6.2	6.9
20	2.0	6.2	13.7	1.5	3.4	8.5	4.5	8.8	11.3	2.7	4.9	6.7	6.9	9.3	11.0	5.0	6.2	6.9
25	1.8	6.1	12.7	0.9	3.2	7.9	4.2	8.4	11.4	2.2	4.4	7.0	6.7	9.4	11.1	5.0	5.7	6.7
30	1.7	6.0	12.3	0.8	3.1	8.1	4.1	7.7	11.0	2.0	4.3	7.0	6.4	8.9	10.9	4.4	5.6	6.6

Notes: (i) 1,000 iterations; (ii) $\hat{M}_1(\hat{p}_0)$, $\hat{M}_1^d(\hat{p}_0)$, the original and modified generalized spectral tests derived under time-varying higher moments, respectively, $\widetilde{M}_1(\hat{p}_0)$, the infeasible original generalized spectral test; (iii) $\bar{p}$, the preliminary lag order used in a plug-in method to choose a data-dependent lag order $\hat{p}_0$. The Parzen kernel is used for $\hat{M}_1(\hat{p}_0)$, $\hat{M}_1^d(\hat{p}_0)$, and $\widetilde{M}_1(\hat{p}_0)$.

more similarly to the infeasible test $\widetilde{M}_1(\hat{p}_0)$ than $\hat{M}_1(\hat{p}_0)$ when $T = 100$, and when $T \geq 250$, $\hat{M}_1^d(\hat{p}_0)$ and $\widetilde{M}_1(\hat{p}_0)$ perform more or less similarly. These results indicate the relative robustness of the modified test $\hat{M}_1^d(\hat{p}_0)$ to parameter estimation uncertainty, illustrating the merits of applying the Wooldridge (1990a) device to the generalized spectral tests.

Next, we turn to the power of the tests $\hat{M}_1(p)$ and $\hat{M}_1^d(p)$. Table 3 reports the empirical rejection rates of the tests at the 5% significance level under DGPs P.1–P.8 using asymptotic and empirical critical values, respectively. For the power using asymptotic critical values, $\hat{M}_1^d(\hat{p}_0)$ is more powerful than or equally powerful to $\hat{M}_1(\hat{p}_0)$ under all scenarios, which is not surprising at all given the fact that $\hat{M}_1(\hat{p}_0)$ tends to underreject under $\mathbb{H}_0$. Under DGPs P.1 (bilinear(1)), P.2 (NMA(1)), P.4 (SETAR(1)) and P.8 (SIGN AR(6)), $\hat{M}_1(\hat{p}_0)$ has substantially lower power than $\hat{M}_1^d(\hat{p}_0)$ when $T = 100$, although its power increases as the sample size T increases and catches up with that of $\hat{M}_1^d(\hat{p}_0)$ when $T = 250$. This is consistent with the substantial underrejection of $\hat{M}_1(\hat{p}_0)$ due to the impact of parameter estimation uncertainty. Under DGPs P.2 (NMA(1)), P.3 (EXP-AR(1)), P.6 (ARMA(1,1)), P.7 (NMA(5)), and P.8 (SIGN AR(6)), both $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$ have unit power except for one scenario when $T = 250$.

We now compare the powers of the tests using empirical critical values, which provide fair comparison of different tests on an equal ground. The empirical critical values are obtained under AR(1)-*i.i.d.* Under DGP P.1 (bilinear(1)), $\hat{M}_1^d(\hat{p}_0)$ is a bit more powerful than $\hat{M}_1(\hat{p}_0)$ when $T = 100$. They have equal or similar power when $T = 250$. The power of $\hat{M}_1^d(\hat{p}_0)$ is relatively more robust to the choice of the preliminary lag order $\bar{p}$ than that of $\hat{M}_1(\hat{p}_0)$. Under DGP P.2 (NMA(1)), $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_1(\hat{p}_0)$ have roughly equal power. The power of $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$ is rather robust to the choice of the preliminary lag order $\bar{p}$ when $T = 250$.

Under DGP P.3 (EXP-AR(1)), $\hat{M}_1^d(\hat{p}_0)$ is slightly less powerful than $\hat{M}_1(\hat{p}_0)$ when $T = 100$. When $T = 250$, $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$ have the same unit power. Also, the power of both tests is robust to the choice of preliminary lag order $\bar{p}$. Under DGP P.4 (SETAR(1)), $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_1(\hat{p}_0)$ also have roughly equal powers for each sample size T.

Under DGP P.5 (STAR(1)), $\hat{M}_1^d(\hat{p}_0)$ is less powerful than $\hat{M}_1(\hat{p}_0)$, particularly when $T = 100$, but the power of $\hat{M}_1^d(\hat{p}_0)$ increases rapidly as T increases. Under DGP P.6 (ARMA(1,1)), the power of $\hat{M}_1^d(\hat{p}_0)$ is slightly smaller than that of $\hat{M}_1(\hat{p}_0)$, and it quickly increases as T increases. When $T = 250$, $\hat{M}_1^d(\hat{p}_0)$ and $\hat{M}_1(\hat{p}_0)$ are equally powerful, and their powers are close to unity.

Under DGP P.7 (NMA(5)), $\hat{M}_1^d(\hat{p}_0)$ is slightly less powerful than $\hat{M}_1(\hat{p}_0)$ when $T = 100$, but its power quickly catches up with that of $\hat{M}_1(\hat{p}_0)$ when $T = 250$. When $T = 250$, both tests have unit power. Under DGP P.8 (SIGN AR(6)), both $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$ are equally powerful for both sample sizes, and they have unit power when $T = 250$.

We note that there exist substantial or significant differences between the power of $\hat{M}_1(\hat{p}_0)$ using the asymptotic critical value and the power of $\hat{M}_1(\hat{p}_0)$

TABLE 3. Empirical powers of tests

T	$T = 100$				$T = 250$				$T = 100$				$T = 250$			
	ACV		ECV		ACV		ECV		ACV		ECV		ACV		ECV	
$\bar{p}$	M_1	M_1^d	M_1	M_1^d	M_1	M_1^d	M_1	M_1^d	M_1	M_1^d	M_1	M_1^d	M_1	M_1^d	M_1	M_1^d
	DGP P.1: Bilinear(1)								DGP P.2: NMA(1)							
10	25.4	49.2	42.6	46.4	57.8	72.8	71.6	71.8	45.2	69.2	66.4	65.4	98.0	98.4	99.2	98.4
15	24.6	48.0	41.0	44.8	57.8	72.8	71.6	71.8	44.0	68.2	64.8	65.0	98.0	98.4	99.2	98.4
20	22.6	45.6	39.0	45.2	57.2	72.8	71.8	71.6	41.0	65.4	63.0	64.8	97.6	98.4	99.0	98.4
25	21.4	44.0	35.8	45.2	54.4	71.4	67.6	70.4	38.0	59.6	58.6	61.8	97.0	98.2	99.0	98.2
30	20.8	42.4	32.6	43.0	52.0	70.0	64.6	69.6	35.6	55.4	57.0	56.0	96.2	98.2	99.0	98.2
	DGP P.3: EXP-AR(1)								DGP P.4: SETAR(1)							
10	89.6	90.4	94.0	90.0	100.0	100.0	100.0	100.0	31.4	51.0	48.0	47.8	81.8	91.8	89.8	90.4
15	89.6	90.2	93.6	90.0	100.0	100.0	100.0	100.0	30.4	50.2	47.2	47.0	81.8	91.8	89.8	90.4
20	89.2	90.2	93.4	90.2	100.0	100.0	100.0	100.0	25.8	45.4	44.0	44.8	80.8	90.8	89.6	90.2
25	88.6	89.0	93.0	90.2	100.0	100.0	100.0	100.0	24.2	41.8	41.8	42.4	78.4	90.0	88.2	89.6
30	88.2	89.0	91.8	89.0	100.0	100.0	100.0	100.0	22.2	38.8	38.8	40.0	75.6	88.2	87.0	87.4
	DGP P.5: STAR(1)								DGP P.6: ARMA(1,1)							
10	35.8	37.8	53.6	35.2	87.2	88.6	93.6	87.2	67.6	79.8	81.2	77.4	99.2	99.8	99.8	99.8
15	35.2	37.4	52.6	34.6	87.2	88.6	93.6	87.2	67.8	79.0	80.0	76.6	99.2	99.8	99.8	99.8
20	35.4	36.4	51.2	36.0	87.0	88.8	93.6	87.2	67.4	76.6	78.4	76.4	99.2	99.6	99.8	99.6
25	34.2	35.0	47.6	36.0	86.4	86.8	93.2	86.4	64.8	74.6	76.2	75.4	99.2	99.6	99.4	99.6
30	32.4	33.4	46.6	33.6	85.0	86.4	92.8	85.8	62.8	73.6	74.4	73.8	99.0	99.4	99.4	99.4
	DGP P.7: NMA(5)								DGP P.8: SIGN AR(6)							
10	82.2	89.6	92.2	88.4	100.0	100.0	100.0	100.0	25.0	42.2	38.8	40.0	83.0	95.2	92.6	94.6
15	82.2	89.4	91.0	87.6	100.0	100.0	100.0	100.0	41.6	58.0	55.4	55.6	99.8	100.0	99.8	99.8
20	80.4	87.0	89.4	87.0	100.0	100.0	100.0	100.0	56.6	73.6	72.0	73.2	100.0	100.0	100.0	100.0
25	76.6	85.0	87.4	86.0	100.0	100.0	100.0	100.0	70.2	84.0	84.8	73.2	100.0	100.0	100.0	100.0
30	74.0	83.4	86.0	83.8	100.0	100.0	100.0	100.0	81.4	90.8	91.2	91.0	100.0	100.0	100.0	100.0

Notes: (i) 500 iterations, 5% significance level; (ii) ACV, the asymptotic critical value; ECV, the empirical critical value; (iii) $\hat{M}_1(\hat{p}_0), \hat{M}_1^d(\hat{p}_0)$, the original and modified generalized spectral tests, respectively; (iv) $\bar{p}$, the preliminary lag order used in a plug-in method to choose a data-dependent lag order $\hat{p}_0$. The Parzen kernel is used for both $\hat{M}_1(\hat{p}_0)$ and $\hat{M}_1^d(\hat{p}_0)$; (v) DGP P.1, $Y_t = 0.5Y_{t-1} + 0.6Y_{t-1}\varepsilon_{t-1} + \varepsilon_t$; DGP P.2, $Y_t = 0.5Y_{t-1} - 0.6e_{t-1}^2 + \varepsilon_t$; DGP P.3, $Y_t = 0.5Y_{t-1} + 10Y_{t-1}\exp(-Y_{t-1}^2) + \varepsilon_t$; DGP P.4, $Y_t = 0.5Y_{t-1}\mathbf{1}(Y_{t-1} \le 0) - 0.5Y_{t-1}\mathbf{1}(Y_{t-1} > 0) + \varepsilon_t$; DGP P.5, $Y_t = 1 - 0.5Y_{t-1} + (-4 - 0.4Y_{t-1})G(2Y_{t-1}) + \varepsilon_t$, 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 $\varepsilon_t \sim i.i.d.N(0,1)$.

using the empirical critical value. In particular, the power of $\hat{M}_1(\hat{p}_0)$ using asymptotic critical value is lower than that of $\hat{M}_1(\hat{p}_0)$ using empirical critical value. In contrast, the power of $\hat{M}_1^d(\hat{p}_0)$ using asymptotic critical value is similar to that of $\hat{M}_1^d(\hat{p}_0)$ using empirical critical value in all cases.

In summary, we have observed the following stylized facts.

- The empirical sizes of the original $\hat{M}_1(\hat{p}_0)$ test are substantially lower than the nominal significance levels, because of the impact of parameter estimation uncertainty. The underrejection is more severe under homoskedastic errors and with more estimated autoregressive parameters. On the other hand, the Wooldridge (1990a) device can effectively reduce the impact of parameter estimation uncertainty; the empirical sizes of the modified $\hat{M}_1^d(\hat{p}_0)$ test are reasonable in most cases, especially when the sample size becomes moderately large. Its sizes are slightly larger than the corresponding significance levels when the preliminary lag order $\bar{p}$ is small, but they improve as both the preliminary lag order $\bar{p}$ and the sample size T increase.
- As the lag order d of the AR(d) model increases, i.e., the number of estimated autoregressive parameters increases, $\hat{M}_1(\hat{p}_0)$ shows more severe underrejections. In contrast, $\hat{M}_1^d(\hat{p}_0)$ is relatively robust to parameter estimation uncertainty and the number of estimated autoregressive parameters in most cases.
- For the power using asymptotic critical values, $\hat{M}_1^d(\hat{p}_0)$ is always more powerful than $\hat{M}_1(\hat{p}_0)$, and the power of $\hat{M}_1^d(\hat{p}_0)$ is more robust to the choice of the preliminary lag order $\bar{p}$ than that of $\hat{M}_1(\hat{p}_0)$ in many cases.
- When using empirical critical values, the $\hat{M}_1^d(\hat{p}_0)$ test is not always as powerful as the $\hat{M}_1(\hat{p}_0)$ test, but it has similar power to $\hat{M}_1(\hat{p}_0)$ in most cases, particularly when the sample size is moderately large. There is not much power loss for the modified test, although its size has been significantly improved. Both tests have omnibus power against all eight DGPs provided the sample size is sufficiently large. They require no knowledge of the lag structure of the potential alternative.
- There exist some substantial differences between the power of $\hat{M}_1(\hat{p}_0)$ using the asymptotic critical value and the power of $\hat{M}_1(\hat{p}_0)$ using the empirical critical value. In particular, the power of $\hat{M}_1(\hat{p}_0)$ using asymptotic critical values is always lower than the power of $\hat{M}_1(\hat{p}_0)$ using empirical critical values, sometime rather substantially. In contrast, the powers of $\hat{M}_1^d(\hat{p}_0)$ using asymptotic and empirical critical values are more or less similar.

6. CONCLUSION

Adapting the Wooldridge (1990a) device to a generalized spectral derivative approach, we propose an improved version of 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. Like the original generalized spectral derivative tests, the modified tests can detect a wide range of model misspecification in mean while being robust to conditional heteroskedasticity and time-varying higher order 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 most appealing feature of the modified tests is that their finite-sample distribution is relatively robust to parameter estimation uncertainty and this is achieved without suffering from significant power loss, as is confirmed in our simulation. These results indicate that the proposed modified tests can be a useful tool in the specification analysis for time series conditional mean models.

NOTES

1. We conjecture that our modified generalized spectral derivative tests are also applicable to threshold autoregressive models with unknown thresholds, but our regularity conditions given in Section 3 rule out this class of models.

2. A potentially useful application is the investigation of possible nonlinear business cycles by $f(\omega,u,v)$. It has been well known that business cycles exhibit asymmetric features, typically with longer expansions and short recessions (e.g., Hamilton, 1989; Diebold and Rudebusch, 1990). The power spectrum, 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 business cycles.

3. The generalized spectrum $f(\omega,u,v)$ is suitable for testing the i.i.d. hypothesis for $\{\varepsilon_t\}$ as is considered in Hong and Lee (2003). It is not suitable for testing $\mathbb{H}_0$, because $\{\varepsilon_t\}$ can be an m.d.s. but not an i.i.d. sequence.

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

5. The use of $E(\varepsilon_t|\varepsilon_{t-j})$ or $\sigma_j^{(1,0)}(0,v)$ for testing $\mathbb{H}_0$ is analogous in spirit to the nonparametric additive models in the nonparametric estimation literature (e.g., Kim and Linton, 2004).

6. We note that 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 an m.d.s. but has $E(\varepsilon_t|\varepsilon_{t-j}) = 0$ *a.s.* for all $j > 0$ is a nonlinear moving-average process $\varepsilon_t = \alpha z_{t-2}z_{t-3} + z_t$, $\{z_t\} \sim i.i.d.(0,\sigma^2)$.

7. For example, consider an MA(1) model: $Y_t = \theta_0\varepsilon_{t-1} + \varepsilon_t = \sum_{j=1}^{\infty} -(-\theta_0)^j Y_{t-j} + \varepsilon_t$. Here, the infeasible information set $I_{t-1} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_1, Y_0, \ldots\}$ contains the entire past history $\{Y_s, s < t\}$ dating back to the infinite past. On the other hand, $I_{t-1}^{\dagger} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_1, \bar{\varepsilon}_0\}$, where $\bar{\varepsilon}_0$ is an assumed initial value for ε_0.

8. Hong and Lee (2005) also considered two other classes of generalized spectral tests, derived under conditional homoskedasticity and i.i.d. regression errors, respectively. These tests can also be modified in the same way as we do for $\hat{M}_1(p)$ to remove the impact of parameter estimation uncertainty.

9. As one referee points out, the asymptotic $N(0,1)$ distribution of $\widetilde{M}_1(p)$ or $\widetilde{M}_1^d(p)$ can be viewed as the normal approximation for a chi-square distribution with the degree of freedom growing to infinity. Because the chi-square distribution is skewed to the right, the normal approximation may be not accurate unless the degree of freedom is sufficiently large. To improve the finite-sample performance of $\widetilde{M}_1(p)$ or $\widetilde{M}_1^d(p)$, one may consider (e.g.) the Chen and Deo (2004) power transformation, which can alleviate the skewness problem. We leave this possibility to future work.

10. We are grateful to a referee for suggesting this alternative approach.

11. We impose $b > 3$ to simplify the proof of Theorem 1. In other words, the condition of $b > 3$ is sufficient but may not be necessary for $\hat{M}_1^d(p)$ to be asymptotically $N(0,1)$ under $\mathbb{H}_0$.

12. Alternatively, one could use the OLS estimator $\hat{\beta}_j(v) = (\sum_{t=j+1}^{T} \hat{G}_t \hat{G}_t')^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v)$ for $0 < j < T$. The asymptotic $N(0,1)$ distribution of $\hat{M}_1^d(p)$ remains unchanged, but the formal proof is more tedious.

13. We do not include exogenous variables as regressors, which is common in practice, because the inclusion of exogenous variables will not have much adverse impact on the size even for the $\hat{M}_1(p)$ test. This is because the estimated parameters corresponding to exogenous variables do not have significant impact on the distribution of the $\hat{M}_1(p)$ test. In contrast, the lagged dependent variables have significant impact on the size of the original test $M_1(p)$, on which we focus in this simulation study.

14. We also consider a generalized autoregressive conditionally heteroskedastic (GARCH) process with an infinite unconditional fourth-order moment. The size performance of the generalized spectral derivative tests is similar.

15. For the Bartlett kernel, the range of the means of $\hat{p}_0$ is wider, from 6.3 to 17.5 when $T = 100$, from 6.4 to 17.8 for $T = 250$, and from 6.6 to 18.0 when $T = 500$. This is apparently due to different bandwidth rules: for the Bartlett kernel, $\hat{p}_0 = \hat{c}_B T^{1/3}$, and for the Parzen kernel, $\hat{p}_0 = \hat{c}_P T^{1/5}$. Nevertheless, the sizes of the tests using the Bartlett kernel are similar to the sizes of the tests using the Parzen kernel. These results are available from the authors upon request.

REFERENCES

Adda, J. & R. Cooper (2003) *Dynamic Economics: Quantitative Methods and Applications*. MIT Press.

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

Brillinger, D. (1965) An introduction to polyspectra. *Annals of Mathematical Statistics* 36, 1351–1374.

Brillinger, D. & M. Rosenblatt (1967a) Asymptotic theory of estimates of *k*-th order spectra. In B. Harris (ed.), *Spectral Analysis of Time Series*, pp. 153–188. Wiley.

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

Chen, W. & R. Deo (2004) Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society, Series B* 66, 117–130.

Cochrane, J. (2001) *Asset Pricing*. Princeton University Press.

Diebold, F.X. & G.D. Rudebusch (1990) A nonparametric investigation of duration dependence in the American business cycle. *Journal of Political Economy* 98, 596–616.

Engle, R.F., D. Lilien, & R.P. Robins (1987) Estimating time-varying risk premia in the term structure: The ARCH-M model. *Econometrica* 55, 391–407.

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

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

Gonçalves, S. & L. Kilian (2004) Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.

Granger, C.W.J. & T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford University Press.

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. & G. Lin (1996) Stock market volatility and the business cycles. *Journal of Applied Econometrics* 11, 573–593.

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. (2000) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
Harvey, C. & A. Siddique (1999) Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34, 465–487.
Harvey, C. & A. Siddique (2000) Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263–1296.
Hausman, J. (1978) Specification tests in econometrics. *Econometrica* 46, 1251–1271.
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. & T.-H. Lee (2003) Diagnostic checking for adequacy of nonlinear time series models. *Econometric Theory* 19, 1065–1121.
Hong, Y. & Y.-J. Lee (2005) Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Review of Economic Studies* 72, 499–541.
Horowitz, J. (2003) Bootstrap methods for Markov processes. *Econometrica* 71, 1049–1082.
Jondeau, E. & M. Rockinger (2003) Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27, 1699–1737.
Kim, W. & O. Linton (2004) The live methods for generalized additive volatility models. *Econometric Theory* 20, 1094–1139.
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*. Academic Press.
Priestley, M. (1988) *Non-Linear and Non-Stationary Time Series Analysis*. Academic Press.
Sargent, T. & L. Ljungqvist (2002) *Recursive Macroeconomic Theory*. MIT Press.
Schwert, G.W. (1989) Why does stock market volatility change over time? *Journal of Finance* 44, 1115–1153.
Stinchcombe, M. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–324.
Subba Rao, T. & M. Gabr (1984) *Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics 24. Springer-Verlag.
Teräsvirta, T. (1994) Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208–218.
Tong, H. (1990) *Nonlinear Time Series: A Dynamic System Approach*. Clarendon Press.
White, H. (1987) Specification testing in dynamic models. In T. Bewley (ed.), *Advances in Econometrics—Fifth World Congress*, vol. 1, pp. 1–58. Cambridge University Press.
White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press.
White, H. (2001) *Asymptotic Theory for Econometricians*, 2nd ed. Academic Press.
Wooldridge, J. (1990a) A unified approach to robust, regression-based specification tests. *Econometric Theory* 6, 17–43.
Wooldridge, J. (1990b) An encompassing approach to conditional mean tests with applications to testing nonnested hypotheses. *Journal of Eonometrics* 45, 331–350.
Wooldridge, J. (1991) On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* 47, 5–46.

MATHEMATICAL APPENDIX

Throughout the Appendix, we let $C \in (1,\infty)$ denote a generic bounded constant.

Proof of Theorem 1. We shall show that $\hat{M}_1^d(p) - \hat{M}_1(p) \xrightarrow{p} 0$. The asymptotic normality of $\hat{M}_1^d(p)$ then follows immediately from Hong and Lee (2005, Thm. 1) under

Assumptions A.1–A.8. Noting that $\varepsilon_t(\theta) \equiv Y_t - g(I_{t-1},\theta)$ in (2.1), where I_t is the unobservable information set from period t to the infinite past, we write

$$\hat{\varepsilon}_t \equiv Y_t - g(I_{t-1}^{\dagger},\hat{\theta}) = \varepsilon_t(\hat{\theta}) + g(I_{t-1},\hat{\theta}) - g(I_{t-1}^{\dagger},\hat{\theta}).$$

Because $I_{t-1}^{\dagger} \neq I_{t-1}$ generally, we have $\hat{\varepsilon}_t \neq \varepsilon_t(\hat{\theta})$, but Assumption A.3 implies

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

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 $G_t(\theta) \equiv (\partial/\partial\theta)g(I_{t-1},\theta)$. It follows from the Cauchy–Schwarz inequality and Assumptions A.2 and A.4 that

$$\sum_{t=1}^{T} [\varepsilon_t(\hat{\theta}) - \varepsilon_t]^2 \leq T\|\hat{\theta} - \theta_0\|^2 T^{-1} \sum_{t=1}^{T} \sup_{\theta\in\Theta_0} \|G_t(\theta)\|^2 = O_P(1), \tag{A.2}$$

where Θ_0 is a neighborhood containing θ_0, the $p\lim(\hat{\theta})$. Both (A.1) and (A.2) imply

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

Throughout, we put $T_j \equiv T - |j|$. To show that $\hat{M}_1^d(p) - \hat{M}_1(p) \xrightarrow{p} 0$, it suffices to show (i)

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

(ii) $p^{-1/2}[\hat{C}_1^d(p) - \hat{C}_1(p)] = O_P(T^{-1/2})$, and (iii) $p^{-1}[\hat{D}_1^d(p) - \hat{D}_1(p)] \xrightarrow{p} 0$, where $\hat{D}_1(p) \propto p$ as shown in Hong and Lee (2005, proof of Thm. 1). For reasons of space, we focus on the proof of (A.4); the proofs for (ii) and (iii) are straightforward. Note that we need to obtain the convergence rate $O_P(T^{-1/2})$ for $p^{-1/2}[\hat{C}_1^d(p) - \hat{C}_1(p)]$ to ensure that replacing $\hat{C}_1^d(p)$ with $\hat{C}_1(p)$ has asymptotically negligible impact given $p/T \to 0$.

To show (A.4), we decompose

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

where

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

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

where $\mathrm{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\hat{\sigma}_j^{(1,0)}(0,v)^*$ is the complex conjugate of $\hat{\sigma}_j^{(1,0)}(0,v)$. Then (A.4) follows from Theorems A.1 and A.2, which are given subsequently, and $p \to \infty$ as $T \to \infty$.

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

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

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

$$\hat{\gamma}_j^{(1,0)}(0,v) - \hat{\sigma}_j^{(1,0)}(0,v) = iT_j^{-1} \sum_{t=j+1}^{T} \hat{\varepsilon}_t [\hat{h}_{t-j}(v) - \hat{\psi}_{t-j}(v)] = -i\hat{\beta}_j(v)' T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\varepsilon}_t, \tag{A.6}$$

where $\hat{h}_{t-j}(v) \equiv \hat{\psi}_{t-j}(v) - \hat{G}_t' \hat{\beta}_j(v)$, $\hat{\beta}_j(v) \equiv (\sum_{t=1}^{T} \hat{G}_t \hat{G}_t')^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v)$, $\hat{G}_t \equiv (\partial/\partial\theta) g(I_{t-1}^{\dagger}, \hat{\theta})$, $\hat{\psi}_{t-j}(v) \equiv e^{iv\hat{\varepsilon}_{t-j}} - \hat{\varphi}_j(v)$, and $\hat{\varphi}_j(v) = T^{-1} \sum_{t=j+1}^{T} e^{iv\hat{\varepsilon}_{t-j}}$. Noting that $\|\hat{\beta}_j(v)\|^2 \le \lambda_{\min}^{-2}(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t') \|T^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v)\|^2$, where $\lambda_{\min}(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t')$ denotes the minimum eigenvalue of $T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t'$, we have

$$\hat{A}_1 \le \lambda_{\min}^{-2}\left(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t'\right) \times \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left\| T^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 \left\| T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\varepsilon}_t \right\|^2 dW(v). \tag{A.7}$$

We first show that $\lambda_{\min}(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t') \ge c > 0$ with probability approaching 1. Given that $\hat{G}_t - G_t = \hat{G}_t - G_t(\hat{\theta}) + G_t(\hat{\theta}) - G_t$, where $G_t \equiv G_t(\theta_0) = (\partial/\partial\theta) g(I_{t-1}, \theta_0)$, we have

$$\left\| T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t' - T^{-1} \sum_{t=1}^{T} G_t G_t' \right\| \le T^{-1} \sum_{t=1}^{T} \|\hat{G}_t - G_t\|^2 + 2T^{-1} \sum_{t=1}^{T} \|\hat{G}_t - G_t\| \|G_t\| = O_P(T^{-1/2}), \tag{A.8}$$

where $T^{-1} \sum_{t=1}^{T} \|\hat{G}_t - G_t\|^2 \le 2T^{-1} \sum_{t=1}^{T} \|\hat{G}_t - G_t(\hat{\theta})\|^2 + 2T^{-1} \sum_{t=1}^{T} \|G_t(\hat{\theta}) - G_t\|^2 = O_P(T^{-1})$ given Assumptions A.2–A.4 and the mean value theorem for the expansion of $G_t(\hat{\theta}) - G_t$.

On the other hand, given Assumption A.5, $\{G_t G_t'\}$ is a strictly stationary mixing process with mixing coefficient $\alpha(j)$. By a standard α-mixing inequality, and Assumption A.2, we have

$$\left\| T^{-1} \sum_{t=1}^{T} G_t G_t' - E(G_t G_t') \right\|^2 \le CT^{-1} \sum_{j=-\infty}^{\infty} \alpha(j)^{(\nu-1)/\nu} \le CT^{-1}. \tag{A.9}$$

It follows from (A.8) and (A.9) that $T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t' \xrightarrow{p} E(G_t G_t')$, and so $\lambda_{\min}(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t') \xrightarrow{p} \lambda_{\min}(EG_t G_t') \ge c > 0$ given nonsingularity of $E(G_t G_t')$ in Assumption A.2. Thus, to bound the order of magnitude for $\hat{A}_1$, we can focus on the term

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \left\| T^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 \left\| T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\varepsilon}_t \right\|^2 dW(v)$$

$$\leq 2 \left\| T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t \right\|^2 \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left\| T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 dW(v)$$

$$+ 2 \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left\| T^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 \left\| T_j^{-1} \sum_{t=1}^{j} \hat{G}_t \hat{\varepsilon}_t \right\|^2 dW(v)$$

$$\equiv 2\hat{B}_1 + 2\hat{B}_2,$$

where we have used the identity that $\sum_{t=j+1}^{T} \hat{G}_t \hat{\varepsilon}_t = (\sum_{t=1}^{T} - \sum_{t=1}^{j}) \hat{G}_t \hat{\varepsilon}_t$. Theorem A.1 follows from Propositions A.1 and A.2, which are given subsequently, $(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t')^{-1} = O_P(1)$, and $p \to \infty$ as $T \to \infty$.

PROPOSITION A.1. *$\hat{B}_1 = O_P(1)$ and $p^{-1/2}\hat{B}_1 \xrightarrow{p} 0$.*

PROPOSITION A.2. *$\hat{B}_2 = O_P(p^2/T)$ and $p^{-1/2}\hat{B}_2 \xrightarrow{p} 0$.*

Proof of Proposition A.1. Noting that $\hat{\varepsilon}_t - \varepsilon_t = \hat{\varepsilon}_t - \varepsilon_t(\hat{\theta}) + \varepsilon_t(\hat{\theta}) - \varepsilon_t$ and $\hat{G}_t - G_t = \hat{G}_t - G_t(\hat{\theta}) + G_t(\hat{\theta}) - G_t$, we decompose

$$T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t = T^{-1} \sum_{t=1}^{T} [\hat{G}_t - G_t(\hat{\theta})][\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})] + T^{-1} \sum_{t=1}^{T} [\hat{G}_t - G_t(\hat{\theta})][\varepsilon_t(\hat{\theta}) - \varepsilon_t]$$

$$+ T^{-1} \sum_{t=1}^{T} [\hat{G}_t - G_t(\hat{\theta})]\varepsilon_t + T^{-1} \sum_{t=1}^{T} G_t(\hat{\theta})[\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})]$$

$$+ T^{-1} \sum_{t=1}^{T} G_t(\hat{\theta})[\varepsilon_t(\hat{\theta}) - \varepsilon_t] + T^{-1} \sum_{t=1}^{T} [G_t(\hat{\theta}) - G_t]\varepsilon_t + T^{-1} \sum_{t=1}^{T} G_t \varepsilon_t$$

$$= O_P(T^{-1} + T^{-1} + T^{-1/2} + T^{-1/2} + T^{-1/2} + T^{-1/2} + T^{-1/2})$$

$$= O_P(T^{-1/2}), \tag{A.10}$$

given Assumptions A.2–A.4, the Cauchy–Schwarz inequality, the mean value theorem, and Markov's inequality or Chebyshev's inequality. The mean value theorem is used for the second, fifth, and sixth terms in (A.10), and Chebyshev's inequality is used for the last term in (A.10), where $\{G_t \varepsilon_t\}$ is an m.d.s. under $\mathbb{H}_0$.

Next we decompose

$$T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) = T_j^{-1} \sum_{t=j+1}^{T} [\hat{G}_t - G_t(\hat{\theta})]\hat{\psi}_{t-j}(v) + T_j^{-1} \sum_{t=j+1}^{T} [G_t(\hat{\theta}) - G_t]\hat{\psi}_{t-j}(v)$$

$$+ T_j^{-1} \sum_{t=j+1}^{T} G_t[\hat{\psi}_{t-j}(v) - \psi_{t-j}(v)]$$

$$+ T_j^{-1} \sum_{t=j+1}^{T} [G_t \psi_{t-j}(v) - \eta_j(v)] + \eta_j(v)$$

$$\equiv \sum_{d=1}^{5} \hat{H}_{dj}(v), \quad \text{say}, \tag{A.11}$$

where $\eta_j(v) \equiv E[G_t \psi_{t-j}(v)]$. For the first term in (A.11), given $|\hat{\psi}_{t-j}(v)| \leq 2$, we have $|\hat{H}_{1j}(v)| \leq 2T_j^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta_0} \|(\partial/\partial\theta) g(I_{t-1}^{\dagger}, \theta) - (\partial/\partial\theta) g(I_{t-1}, \theta)\|$. It follows from Assumptions A.3, A.6, and A.7, and Markov's inequality that

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j |\hat{H}_{1j}(v)|^2 \, dW(v)$$

$$\leq 4 \int dW(v) \left[\sum_{t=1}^{T} \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger}, \theta) - \frac{\partial}{\partial\theta} g(I_{t-1}, \theta) \right\| \right]^2 \sum_{j=1}^{T-1} k^2(j/p) T_j^{-1}$$

$$= O_P(p/T), \qquad \textbf{(A.12)}$$

where $\sum_{j=1}^{T-1} k^2(j/p) T_j^{-1} = O(p/T)$ as is shown in Hong (1999, (A.15)).

For the second term in (A.11), $\hat{H}_{2j}(v)$, using the mean value theorem and the fact that $|\hat{\psi}_{t-j}(v)| \leq 2$, we have $|\hat{H}_{2j}(v)| \leq 2\|\hat{\theta} - \theta_0\| T_j^{-1} \sum_{t=j+1}^{T} \sup_{\theta \in \Theta_0} \|(\partial^2/\partial\theta\partial\theta') g(I_{t-1}, \theta)\|$. It follows that

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j |\hat{H}_{2j}(v)|^2 \, dW(v)$$

$$\leq 4\|\hat{\theta} - \theta_0\|^2 \int dW(v) \sum_{j=1}^{T-1} k^2(j/p) T_j \left[T_j^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta_0} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} g(I_{t-1}, \theta) \right\| \right]^2$$

$$= O_P(p), \qquad \textbf{(A.13)}$$

given Assumptions A.2, A.4, A.6, and A.7, Minkowski's inequality, and Markov's inequality.

For the third term $\hat{H}_{3j}(v)$, we write

$$\hat{H}_{3j}(v) = T_j^{-1} \sum_{t=j+1}^{T} G_t (e^{iv\hat{\varepsilon}_{t-j}} - e^{iv\varepsilon_{t-j}}) - \left(T_j^{-1} \sum_{t=j+1}^{T} G_t \right) [\hat{\varphi}_j(v) - \varphi(v)]$$

$$\equiv \hat{H}_{31j}(v) - \hat{H}_{32j}(v).$$

For the term $\hat{H}_{31j}(v)$, using the identity that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued z_1 and z_2, we have $\|\hat{H}_{31j}(v)\| \leq |v| T_j^{-1} \sum_{t=j+1}^{T} \|G_t\| |\hat{\varepsilon}_{t-j} - \varepsilon_{t-j}|$. It follows from the Cauchy–Schwarz inequality and (A.3) that

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \|\hat{H}_{31j}(v)\|^2 \, dW(v)$$

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

$$= O_P(p). \qquad \textbf{(A.14)}$$

For the term $\hat{H}_{32j}(v)$, we first decompose

$$|\hat{\varphi}_j(v) - \varphi(v)| \le \left| T_j^{-1} \sum_{t=j+1}^{T} (e^{iv\hat{\varepsilon}_{t-j}} - e^{iv\varepsilon_{t-j}}) \right| + \left| T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(v) \right|$$
$$\le |v| T_j^{-1} \sum_{t=j+1}^{T} |\hat{\varepsilon}_{t-j} - \varepsilon_{t-j}| + \left| T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(v) \right|.$$

It follows from the Cauchy–Schwarz inequality and Markov's inequality that

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \|\hat{H}_{32j}(v)\|^2 \, dW(v)$$
$$\le 2 \int v^2 \, dW(v) \left[\sum_{t=1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 \right] \sum_{j=1}^{T-1} k^2(j/p) \left\| T_j^{-1} \sum_{t=j+1}^{T} G_t \right\|^2$$
$$+ 2 \sum_{j=1}^{T-1} k^2(j/p) T_j \left\| T_j^{-1} \sum_{t=j+1}^{T} G_t \right\|^2 \int \left| T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(v) \right|^2 dW(v)$$
$$= O_P(p) + O_P(p) = O_P(p), \tag{A.15}$$

where we used the facts that $E\|T_j^{-1}\sum_{t=j+1}^{T} G_t\|^2 \le [T_j^{-1}\sum_{t=j+1}^{T}(E\|G_t\|^2)^{1/2}]^2$ and $E|T_j^{-1}\sum_{t=j+1}^{T}\psi_{t-j}(v)|^2 \le CT_j^{-1}$, where the latter follows from Assumption A.5 and a standard mixing inequality.

We now consider the fourth term in (A.11), $\hat{H}_{4j}(v)$. Put $\tilde{G}_t \equiv G_t - EG_t$. We decompose

$$\hat{H}_{4j}(v) = E(G_t) \left[T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(v) \right] + \left[T_j^{-1} \sum_{t=j+1}^{T} \tilde{G}_t \psi_{t-j}(v) - \eta_j(v) \right].$$

By Assumption A.5 and a standard mixing inequality, we have $E|T_j^{-1}\sum_{t=j+1}^{T} \psi_{t-j}(v)|^2 \le CT_j^{-1}$ and $E\|T_j^{-1}\sum_{t=j+1}^{T}[\tilde{G}_t\psi_{t-j}(v) - \eta_j(v)]\|^2 \le CT_j^{-1}$ (see Hong, 1999, (A.7) and related proof). It follows that

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \|\hat{H}_{4j}(v)\|^2 \, dW(v)$$
$$\le 2\|E(G_t)\|^2 \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left| T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(v) \right|^2 dW(v)$$
$$+ 2 \sum_{j=1}^{T-1} k^2(j/p) T_j \int \left\| T_j^{-1} \sum_{t=j+1}^{T} \tilde{G}_t \psi_{t-j}(v) - \eta_j(v) \right\|^2 dW(v)$$
$$= O_P(p) + O_P(p) = O_P(p). \tag{A.16}$$

Finally, for the last term in (A.11), we have

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \|\hat{H}_{5j}(v)\|^2 \, dW(v) \le T \sum_{j=1}^{\infty} \int \|\eta_j(v)\|^2 \, dW(v) = O(T), \tag{A.17}$$

where we have made use of the facts that $\sum_{j=-\infty}^{\infty} \sup_{v\in\mathbb{R}} \|\eta_j(v)\| \le C \sum_{j=-\infty}^{\infty} \alpha(j)^{(\nu-1)/\nu} < \infty$ given Assumption A.5, and $|k(\cdot)| \le 1$ from Assumption A.6.

Collecting (A.10)–(A.17), we obtain

$$\left\| T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t \right\|^2 \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left\| T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 dW(v)$$

$$= O_P(T^{-1}) O_P(p/T + p + p + T) = O_P(1). \qquad \blacksquare$$

Proof of Proposition A.2. We first write

$$\hat{B}_2 \le \left[\int \sum_{j=1}^{T-1} |k(j/p)| T_j \left\| T^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 dW(v) \right]$$

$$\times \left[\sum_{j=1}^{T-1} |k(j/p)| \left\| T_j^{-1} \sum_{t=1}^{j} \hat{G}_t \hat{\varepsilon}_t \right\|^2 \right]. \qquad \textbf{(A.18)}$$

For the term in the first bracket in (A.18), we can obtain

$$\int \sum_{j=1}^{T-1} |k(j/p)| T_j \left\| T^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right\|^2 dW(v) = O_P(T), \qquad \textbf{(A.19)}$$

by analogous reasoning to (A.11)–(A.17). Note that the difference of having the factor $|k(\cdot)|$ rather than the factor $k^2(\cdot)$ does not change the order of magnitude for the term in (A.19).

Next we consider the term in the second bracket in (A.18). As in (A.10), we decompose

$$T_j^{-1} \sum_{t=1}^{j} \hat{G}_t \hat{\varepsilon}_t = T_j^{-1} \sum_{t=1}^{j} [\hat{G}_t - G_t(\hat{\theta})][\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})] + T_j^{-1} \sum_{t=1}^{j} [\hat{G}_t - G_t(\hat{\theta})][\varepsilon_t(\hat{\theta}) - \varepsilon_t]$$

$$+ T_j^{-1} \sum_{t=1}^{j} [\hat{G}_t - G_t(\hat{\theta})]\varepsilon_t + T_j^{-1} \sum_{t=1}^{j} G_t(\hat{\theta})[\hat{\varepsilon}_t - \varepsilon_t(\hat{\theta})]$$

$$+ T_j^{-1} \sum_{t=1}^{j} G_t(\hat{\theta})[\varepsilon_t(\hat{\theta}) - \varepsilon_t] + T_j^{-1} \sum_{t=1}^{j} [G_t(\hat{\theta}) - G_t]\varepsilon_t + T_j^{-1} \sum_{t=1}^{j} G_t \varepsilon_t$$

$$\equiv \sum_{d=1}^{7} \hat{C}_{dj}, \quad \text{say.} \qquad \textbf{(A.20)}$$

For the first term $\hat{C}_{1j}$ in (A.20), we have

$$|\hat{C}_{1j}| \le T_j^{-1} \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta_0} |g(I_{t-1}^{\dagger}, \theta) - g(I_{t-1}, \theta)|^2 \right]^{1/2}$$

$$\times \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger}, \theta) - \frac{\partial}{\partial\theta} g(I_{t-1}, \theta) \right\|^2 \right]^{1/2},$$

by the Cauchy–Schwarz inequality. It follows that

$$\begin{aligned}
&\sum_{j=1}^{T-1} |k(j/p)||\hat{C}_{1j}|^2 \\
&\quad \le \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta_0} |g(I_{t-1}^{\dagger},\theta) - g(I_{t-1},\theta)|^2 \right] \\
&\qquad \times \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger},\theta) - \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right] \sum_{j=1}^{T-1} |k(j/p)| T_j^{-2} \\
&\quad = O_P(p/T^2), \qquad \textbf{(A.21)}
\end{aligned}$$

where $\sum_{j=1}^{T-1}|k(j/p)|T_j^{-2} = O(p/T^2)$, following analogous reasoning to (A.15) of Hong (1999).

Next, we consider $\hat{C}_{2j}$. By the mean value theorem and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
|\hat{C}_{2j}| &\le \|\hat{\theta} - \theta_0\| T_j^{-1} \sum_{t=1}^{j} \|\hat{G}_t - G_t(\hat{\theta})\| \|G_t(\bar{\theta})\| \\
&\le \|\hat{\theta} - \theta_0\| \left[T_j^{-1} \sum_{t=1}^{j} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger},\theta) - \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right]^{1/2} \\
&\quad \times \left[T_j^{-1} \sum_{t=1}^{j} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right]^{1/2}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{j=1}^{T-1} |k(j/p)||\hat{C}_{2j}|^2 \\
&\quad \le \|\hat{\theta} - \theta_0\|^2 \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger},\theta) - \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right] \\
&\qquad \times \left[\sum_{j=1}^{T-1} |k(j/p)| T_j^{-2} \sum_{t=1}^{j} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right] \\
&\quad = O_P(T^{-1}) O_P(1) O_P(p^2/T^2) = O_P(p^2/T^3), \qquad \textbf{(A.22)}
\end{aligned}$$

given Assumptions A.2–A.4 and A.6, where the term in the second bracket is $O_P(p^2/T^2)$ by Markov's inequality and the fact that

$$\begin{aligned}
&\sum_{j=1}^{T-1} |k(j/p)| T_j^{-2} \sum_{t=1}^{j} E \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \\
&\quad \le C p^2 \left[p^{-1} \sum_{j=1}^{T-1} (j/p) |k(j/p)| T_j^{-2} \right] = O(p^2/T^2),
\end{aligned}$$

where $p^{-1}\sum_{j=1}^{T-1}(j/p)|k(j/p)|T_j^{-2} \to T^{-2}\int_0^\infty z|k(z)|\,dz$ as $p \to \infty$ and $T \to \infty$.

Next, for the $\hat{C}_{3j}$ term in (A.20), by the Cauchy–Schwarz inequality, we have

$$\|\hat{C}_{3j}\| \leq \left[T_j^{-1} \sum_{t=1}^{T} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger},\theta) - \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right]^{1/2} \left(T_j^{-1} \sum_{t=1}^{j} \varepsilon_t^2 \right)^{1/2}.$$

It follows from Assumptions A.2, A.3, A.5, and A.6 and Markov's inequality that

$$\begin{aligned}\sum_{j=1}^{T-1} |k(j/p)| \|\hat{C}_{3j}\|^2 &\leq \left[\sum_{t=1}^{T} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1}^{\dagger},\theta) - \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right] \\ &\quad \times \left[\sum_{j=1}^{T-1} |k(j/p)| T_j^{-2} \sum_{t=1}^{j} \varepsilon_t^2 \right] \\ &= O_P(1) O_P(p^2/T^2) = O_P(p^2/T^2). \qquad \textbf{(A.23)}\end{aligned}$$

Similarly, following analogous reasoning to $\hat{C}_{3j}$, we can obtain

$$\begin{aligned}\sum_{j=1}^{T-1} |k(j/p)| \|\hat{C}_{4j}\|^2 &\leq \left\{ \sum_{t=1}^{T} \sup_{\theta\in\Theta_0} [g(I_{t-1}^{\dagger},\theta) - g(I_{t-1},\theta)]^2 \right\} \\ &\quad \times \sum_{j=1}^{T-1} |k(j/p)| T_j^{-2} \sum_{t=1}^{j} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \\ &= O_P(p^2/T^2). \qquad \textbf{(A.24)}\end{aligned}$$

Next, using the mean value theorem for $\varepsilon_t(\hat{\theta}) - \varepsilon_t$ and $G_t(\hat{\theta}) - G_t$, respectively, we can obtain

$$\begin{aligned}\sum_{j=1}^{T-1} |k(j/p)| \|\hat{C}_{5j}\|^2 &\leq \|\hat{\theta} - \theta_0\|^2 \sum_{j=1}^{T-1} |k(j/p)| \left[T_j^{-1} \sum_{t=1}^{j} \sup_{\theta\in\Theta_0} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} g(I_{t-1},\theta) \right\|^2 \right]^2 \\ &= O_P(p^3/T^3) \qquad \textbf{(A.25)}\end{aligned}$$

and

$$\begin{aligned}\sum_{j=1}^{T-1} |k(j/p)| \|\hat{C}_{6j}\|^2 &\leq \|\hat{\theta} - \theta_0\|^2 \sum_{j=1}^{T-1} |k(j/p)| \\ &\quad \times \left[T_j^{-1} \sum_{t=1}^{j} \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} g(I_{t-1},\theta) \right\|^2 \right] \left(T_j^{-1} \sum_{t=1}^{j} \varepsilon_t^2 \right) \\ &= O_P(p^3/T^3), \qquad \textbf{(A.26)}\end{aligned}$$

given Assumptions A.2 and A.4–A.6, where we have made use of the fact that $p^{-1}\sum_{j=1}^{T-1}(j/p)^2|k(j/p)|T_j^{-2} = O(T^{-2})$ given Assumption A.6.

Finally, for the $\hat{C}_{7j}$ term in (A.20), by Markov's inequality and the fact that $E\|T_j^{-1}\sum_{t=1}^{j} G_t\varepsilon_t\|^2 \leq (j/T_j^2)(E\|G_t\|^4)^{1/2}[E(\varepsilon_t^4)]^{1/2}$ given the m.d.s. property of $\{\varepsilon_t\}$ under $\mathbb{H}_0$, we have

$$\sum_{j=1}^{T-1} |k(j/p)| \|\hat{C}_{7j}\|^2 = \sum_{j=1}^{T-1} |k(j/p)| \left\| T_j^{-1} \sum_{t=1}^{j} G_t \varepsilon_t \right\|^2 = O_P(p^2/T^2). \quad \textbf{(A.27)}$$

Collecting (A.20)–(A.27) and $p/T \to 0$, we obtain

$$\sum_{j=1}^{T-1} |k(j/p)| \left\| T_j^{-1} \sum_{t=1}^{j} \hat{G}_t \hat{\varepsilon}_t \right\|^2 = O_P(p^2/T^2). \quad \textbf{(A.28)}$$

It follows from (A.18), (A.19), and (A.28) that $\hat{B}_2 = O_P(T)O_P(p^2/T^2) = o_P(p^{1/2})$, given $p^3/T^2 \to 0$. This completes the proof of Theorem A.1. ■

Proof of Theorem A.2. Given (A.6), we can write

$$\begin{aligned}
&-[\hat{\gamma}_j^{(1,0)}(0,v) - \hat{\sigma}_j^{(1,0)}(0,v)]\, \hat{\sigma}_j^{(1,0)}(0,v)^* \\
&\quad = i\hat{\beta}_j(v)' \left(T_j^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t \right) \hat{\sigma}_j^{(1,0)}(0,v)^* - i\hat{\beta}_j(v)' \left(T_j^{-1} \sum_{t=1}^{j} \hat{G}_t \hat{\varepsilon}_t \right) \hat{\sigma}_j^{(1,0)}(0,v)^* \\
&\quad \equiv i\hat{B}_{1j}(v) - i\hat{B}_{2j}(v). \qquad \textbf{(A.29)}
\end{aligned}$$

First we consider the term $\hat{B}_{1j}(v)$ in (A.29). Recalling that $\hat{\beta}_j(v) = (\sum_{t=1}^{T} \hat{G}_t \hat{G}_t')^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v)$, we have

$$\begin{aligned}
&\int \sum_{j=1}^{T-1} k^2(j/p) T_j \hat{B}_{1j}(v)\, dW(v) \\
&\quad = \left(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t \right) \left(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t' \right)^{-1} \\
&\qquad \times \int \sum_{j=1}^{T-1} k^2(j/p) T_j \left(T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right) \hat{\sigma}_j^{(1,0)}(0,v)^*\, dW(v).
\end{aligned}$$

As shown in the proof of Theorem A.1, we have $(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t')^{-1} = O_P(1)$ and $T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t = O_P(T^{-1/2})$ in (A.10). Therefore, we only need to bound the order of magnitude for the term

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \left(T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right) \hat{\sigma}_j^{(1,0)}(0,v)^*\, dW(v).$$

Using (A.11), we can decompose

$$\begin{aligned}
&\int \sum_{j=1}^{T-1} k^2(j/p) T_j \left(T_j^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v) \right) \hat{\sigma}_j^{(1,0)}(0,v)^*\, dW(v) \\
&\quad = \sum_{d=1}^{5} \int \sum_{j=1}^{T-1} k^2(j/p) T_j \hat{H}_{dj}(v) \hat{\sigma}_j^{(1,0)}(0,v)^*\, dW(v). \qquad \textbf{(A.30)}
\end{aligned}$$

For the first term in (A.30), $\hat{H}_{1j}(v)$, we have

$$\begin{aligned}
&\left\| \sum_{j=1}^{T-1} k^2(j/p)T_j \int \hat{H}_{1j}(v)\hat{\sigma}_j^{(1,0)}(0,v)^*\,dW(v) \right\| \\
&\quad \le \left[\sum_{j=1}^{T-1} k^2(j/p)T_j \int \|\hat{H}_{1j}(v)\|^2\,dW(v) \right]^{1/2} \\
&\qquad \times \left[\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{\sigma}_j^{(1,0)}(0,v)|^2\,dW(v) \right]^{1/2} \\
&\quad = O_P(p^{1/2}/T^{1/2})O_P(p^{1/2}) = O_P(p/T^{1/2}), \qquad \textbf{(A.31)}
\end{aligned}$$

by the Cauchy–Schwarz inequality, (A.12), and the fact that

$$\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{\sigma}_j^{(1,0)}(0,v)|^2\,dW(v) = O_P(p) \qquad \textbf{(A.32)}$$

under $\mathbb{H}_0$, as implied by Hong and Lee (2005, Thm. 1).

Similarly, for the second to fourth terms $\hat{H}_{2j}(v)$, $\hat{H}_{3j}(v)$, and $\hat{H}_{4j}(v)$, by the Cauchy–Schwarz inequality and (A.13)–(A.16), we have for $d = 2,3,4$,

$$\begin{aligned}
&\left\| \int \sum_{j=1}^{T-1} k^2(j/p)T_j \hat{H}_{dj}(v)\hat{\sigma}_j^{(1,0)}(0,v)^*\,dW(v) \right\| \\
&\quad \le \left[\sum_{j=1}^{T-1} k^2(j/p)T_j \int \|\hat{H}_{dj}(v)\|^2\,dW(v) \right]^{1/2} \\
&\qquad \times \left[\sum_{j=1}^{T-1} k^2(j/p)T_j \int |\hat{\sigma}_j^{(1,0)}(0,v)|^2\,dW(v) \right]^{1/2} \\
&\quad = O_P(p^{1/2})O_P(p^{1/2}) = O_P(p). \qquad \textbf{(A.33)}
\end{aligned}$$

Finally, we consider the last term $\hat{H}_{5j}(v)$. Recalling that $\hat{H}_{5j}(v) = \eta_j(v)$, and using the triangle inequality, we have

$$\begin{aligned}
&\left\| \int \sum_{j=1}^{T-1} k^2(j/p)T_j \hat{H}_{5j}(v)\hat{\sigma}_j^{(1,0)}(0,v)^*\,dW(v) \right\| \\
&\quad \le \left\| \int \sum_{j=1}^{T-1} k^2(j/p)T_j \eta_j(v)\tilde{\sigma}_j^{(1,0)}(0,v)^*\,dW(v) \right\| \\
&\qquad + \left\| \int \sum_{j=1}^{T-1} k^2(j/p)T_j \eta_j(v)[\hat{\sigma}_j^{(1,0)}(0,v)^* - \tilde{\sigma}_j^{(1,0)}(0,v)^*]\,dW(v) \right\| \\
&\quad = O_P(T^{1/2}) + O_P(T^{1/2})O_P(1) = O_P(T^{1/2}), \qquad \textbf{(A.34)}
\end{aligned}$$

where the first term is $O_P(T^{1/2})$ by Markov's inequality, Minkowski's inequality, $\sum_{j=-\infty}^{\infty} \sup_{v\in\mathbb{R}} \|\eta_j(v)\| < \infty$, $|k(\cdot)| \le 1$, and $\sup_{v\in\mathbb{R}} E|\tilde{\sigma}_j^{(1,0)}(0,v)|^2 \le CT_j^{-1}$ under $\mathbb{H}_0$; the second term is also $O_P(T^{1/2})$ by the Cauchy–Schwarz inequality,

$\sum_{j=-\infty}^{\infty} \sup_{v\in\mathbb{R}} \|\eta_j(v)\| < \infty$, $|k(\cdot)| \le 1$, and $\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\hat{\sigma}_j^{(1,0)}(0,v) - \tilde{\sigma}_j^{(1,0)}(0,v)|^2 \, dW(v) = O_P(1)$ as is shown in Hong and Lee (2005, Prop. A.1).

Combining (A.30)–(A.34), $(T^{-1}\sum_{t=1}^{T} \hat{G}_t \hat{G}_t)^{-1} = O_P(1)$, and $T^{-1}\sum_{t=1}^{T} \hat{G}_t \hat{\varepsilon}_t = O_P(T^{-1/2})$, we obtain

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j \hat{B}_{1j}(v)\, dW(v) = O_P(T^{-1/2}) O_P(1) O_P(T^{1/2}) = O_P(1). \tag{A.35}$$

It remains to consider the term $\hat{B}_{2j}(v)$ in (A.29). By the Cauchy–Schwarz inequality and $\hat{B}_2 = O_P(p^2/T)$ from Proposition A.2, we have

$$\begin{aligned} &\left\| \int \sum_{j=1}^{T-1} k^2(j/p) T_j \hat{B}_{2j}(v)\, dW(v) \right\| \\ &\quad \le \lambda_{\min}^{-1}\left(T^{-1}\sum_{t=1}^{T} \hat{G}_t \hat{G}_t'\right)(\hat{B}_2)^{1/2} \left[\sum_{j=1}^{T-1} k^2(j/p) T_j |\hat{\sigma}_j^{(1,0)}(0,v)|^2\, dW(v)\right]^{1/2} \\ &\quad = O_P(1) O_P(p/T^{1/2}) O_P(p^{1/2}) = O_P(p^{3/2}/T^{1/2}). \end{aligned} \tag{A.36}$$

The desired result of Theorem A.2 then follows from (A.29), (A.35), (A.36), and $p^2/T \to 0$. ■

Proof of Theorem 2. Recall that $h_{t-j}(v) = \psi_{t-j}(v) - G_t' \beta_j(v)$, where $\psi_{t-j}(v) = e^{iv\varepsilon_{t-j}} - \varphi(v)$, $G_t = (\partial/\partial\theta) g(I_{t-1}, \theta_0)$, and $\beta_j(v) = [E(G_t G_t')]^{-1} E[G_t \psi_{t-j}(v)]$. We define the following pseudo test statistic:

$$\overline{M}_1^d(p) = \left[\sum_{j=1}^{T-1} k^2(j/p) T_j \int |\bar{\gamma}_j^{(1,0)}(0,v)|^2\, dW(v) - \bar{C}_1^d(p)\right] \Big/ \sqrt{\bar{D}_1^d(p)},$$

where $\bar{\gamma}_j^{(1,0)}(0,v) = T_j^{-1} \sum_{t=j+1}^{T} i\varepsilon_t h_{t-j}(v)$,

$$\bar{C}_1^d(p) = \sum_{j=1}^{T-1} k^2(j/p) \int T_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t^2 |h_{t-j}(v)|^2\, dW(v),$$

$$\begin{aligned} \bar{D}_1^d(p) &= \sum_{j=1}^{T-2}\sum_{l=1}^{T-2} k^2(j/p) k^2(l/p) \\ &\quad \times \iint \left| \frac{1}{T - \max(j,l)} \sum_{t=\max(j,l)+1}^{T} \varepsilon_t^2 h_{t-j}(v) h_{t-l}(v) \right|^2 dW(u)\, dW(v). \end{aligned}$$

The proof of Theorem 2 consists of the proofs of Theorems A.3 and A.4, which follow, where Theorem A.3 shows that replacing the estimated residuals $\{\hat{\varepsilon}_t\}_{t=1}^{T}$ with the unobservable sample $\{\varepsilon_t\}_{t=1}^{T}$ and replacing the OLS estimators $\{\hat{\beta}_j(v)\}_{j=1}^{T-1}$ with their population counterparts do not affect the asymptotic behavior of $(p^{1/2}/T)\hat{M}_1^d(p)$ under $\mathbb{H}_A$. Theorem A.4 shows that $(p^{1/2}/T)\overline{M}_1^d(p)$ converges to a well-defined probability limit under $\mathbb{H}_A$ from which the $\hat{M}_1^d(p)$ test gains its power.

THEOREM A.3. *Under the conditions of Theorem 2,* $(p^{1/2}/T)[\hat{M}_1^d(p) - \overline{M}_1^d(p)] \xrightarrow{p} 0$.

THEOREM A.4. *Under the conditions of Theorem 2,*

$$(p^{1/2}/T)\overline{M}_1^d(p) \xrightarrow{p} \left[2D_d \int_0^\infty k^2(z)\,dz\right]^{-1/2} \times \pi \iint_{-\pi}^{\pi} |S^{(0,1,0)}(\omega,0,v) - S_0^{(0,1,0)}(\omega,0,v)|^2\, d\omega dW(v).$$

Proof of Theorem A.3. It suffices to show that (i)

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/p) T_j [|\hat{\gamma}_j^{(1,0)}(0,v)|^2 - |\overline{\gamma}_j^{(1,0)}(0,v)|^2]\, dW(v) \xrightarrow{p} 0; \tag{A.37}$$

(ii) $p^{-1}[\hat{C}_1^d(p) - \bar{C}_1^d(p)] = O_P(1)$; (iii) $p^{-1}[\hat{D}_1^d(p) - \overline{D}_1^d(p)] \xrightarrow{p} 0$; and (iv) $\overline{D}_1^d(p) \propto p$. Here we focus on the proof of (i). The proofs of (ii) and (iii) are straightforward, and the result (iv) that $\overline{D}_1^d(p) \propto p$ can be shown in a similar way to the proof of the $\tilde{D}_1(p)$ term in Hong and Lee (2005, proof of Thm. 1).

To show (A.37), we decompose

$$\int \sum_{j=1}^{T-1} k^2(j/p) T_j [|\hat{\gamma}_j^{(1,0)}(0,v)|^2 - |\overline{\gamma}_j^{(1,0)}(0,v)|^2]\, dW(v) = \hat{A}_3 + 2\,\mathrm{Re}(\hat{A}_4), \tag{A.38}$$

where

$$\hat{A}_3 = \int \sum_{j=1}^{T-1} k^2(j/p) T_j |\hat{\gamma}_j^{(1,0)}(0,v) - \overline{\gamma}_j^{(1,0)}(0,v)|^2\, dW(v),$$

$$\hat{A}_4 = \int \sum_{j=1}^{T-1} k^2(j/p) T_j [\hat{\gamma}_j^{(1,0)}(0,v) - \overline{\gamma}_j^{(1,0)}(0,v)] \overline{\gamma}_j^{(1,0)}(0,v)^*\, dW(v).$$

From the Cauchy–Schwarz inequality and the fact that $T^{-1}\int \sum_{j=1}^{T-1} k^2(j/p) T_j |\overline{\gamma}_j^{(1,0)}(0,v)|^2\, dW(v) = O_P(1)$ under $\mathbb{H}_A$ as is implied by Theorem A.4 (the proof of Theorem A.4 does not depend on Theorem A.3) and $\sum_{j=1}^{T-1} k^2(j/p) T_j^{-1} = O(p/T)$, it suffices to show that $T^{-1}\hat{A}_3 \xrightarrow{p} 0$.

By straightforward algebra, we have $j > 0$,

$$\begin{aligned}
&\hat{\gamma}_j^{(1,0)}(0,v) - \overline{\gamma}_j^{(1,0)}(0,v) \\
&= iT_j^{-1} \sum_{t=j+1}^{T} [\hat{\varepsilon}_t \hat{h}_{t-j}(v) - \varepsilon_t h_{t-j}(v)] \\
&= iT_j^{-1} \sum_{t=j+1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)\hat{h}_{t-j}(v) + iT_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t [\hat{h}_{t-j}(v) - h_{t-j}(v)] \\
&= iT_j^{-1} \sum_{t=j+1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)\hat{h}_{t-j}(v) + iT_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t [e^{iv\hat{\varepsilon}_{t-j}} - e^{iv\varepsilon_{t-j}}] \\
&\quad - [\hat{\varphi}_j(v) - \varphi(v)] T_j^{-1} \sum_{t=j+1}^{T} i\varepsilon_t \\
&\quad - iT_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t (\hat{G}_t - G_t)'\hat{\beta}_j(v) - \left(iT_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t G_t'\right)[\hat{\beta}_j(v) - \beta_j(v)] \\
&\equiv i[\hat{B}_{1j}(v) + \hat{B}_{2j}(v) - \hat{B}_{3j}(v) - \hat{B}_{4j}(v) - \hat{B}_{5j}(v)].
\end{aligned} \tag{A.39}$$

For the first term in (A.39), by the Cauchy–Schwarz inequality, we have

$$|\hat{B}_{1j}(v)|^2 \leq \left[T_j^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 \right] \left[T_j^{-1} \sum_{t=j+1}^{T} |\hat{h}_{t-j}(v)|^2 \right] \leq 4T_j^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2,$$

where $T_j^{-1} \sum_{t=j+1}^{T} |\hat{h}_{t-j}(v)|^2 \leq T_j^{-1} \sum_{t=j+1}^{T} |\hat{\psi}_{t-j}(v)|^2$ given the fact that $\hat{h}_{t-j}(v)$ is the OLS estimated residual of regressing $\hat{\psi}_{t-j}(v)$ on $\hat{G}_t$. It follows from (A.3) and $\sum_{j=1}^{T-1} k^2(j/p) = O_P(p)$ that

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

$$\leq \int dW(v) \left[\sum_{t=1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2 \right] T^{-1} \sum_{j=1}^{T-1} k^2(j/p) = O_P(p/T). \qquad \textbf{(A.40)}$$

Next, using the inequality that $|e^{iz_1} - e^{iz_2}| \leq |z_1 - z_2|$ for any real-valued z_1 and z_2, and the Cauchy–Schwarz inequality, we have $|\hat{B}_{2j}(v)|^2 \leq v^2 (T_j^{-1} \sum_{t=1}^{T} \varepsilon_t^2) T_j^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_t - \varepsilon_t)^2$. It follows from (A.3) and Markov's inequality that

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

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

$$= O_P(1) O_P(1) O_P(p/T) = O_P(p/T). \qquad \textbf{(A.41)}$$

For the third term in (A.39), we have

$$|\hat{B}_{3j}(v)|^2 \leq |\hat{\varphi}_j(v) - \varphi(v)|^2 \left(T_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t \right)^2 \leq 4 \left(T_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t \right)^2$$

given $|\hat{\varphi}_j(v)| \leq 1$ and $|\varphi(v)| \leq 1$. It follows that

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

$$\leq 4T^{-1} \sum_{j=1}^{T-1} k^2(j/p) T_j \left(T_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t \right)^2 \int dW(v) = O_P(p/T), \qquad \textbf{(A.42)}$$

by Markov's inequality and the fact that $E(T_j^{-1} \sum_{t=j+1}^{T} \varepsilon_t)^2 \leq C T_j^{-1}$ using Assumption A.5 and a standard mixing inequality. Note that $\{\varepsilon_t\}$ is not an m.d.s. under $\mathbb{H}_A$.

Now, we turn to the fourth term $\hat{B}_{4j}(v)$ in (A.39). Recalling that $\hat{\beta}_j(v) = (\sum_{t=1}^{T} \hat{G}_t \hat{G}_t')^{-1} \sum_{t=j+1}^{T} \hat{G}_t \hat{\psi}_{t-j}(v)$ and $|\hat{\psi}_{t-j}(v)| \leq 2$, we have

$$\max_{1 \leq j < T} \sup_{v \in \mathbb{R}} \|\hat{\beta}_j(v)\| \leq 2\lambda_{\min}^{-1} \left(T^{-1} \sum_{t=1}^{T} \hat{G}_t \hat{G}_t' \right) T^{-1} \sum_{t=j+1}^{T} \|\hat{G}_t\|$$

$$= O_P(1) O_P(1) = O_P(1), \qquad \textbf{(A.43)}$$

where we made use of the fact that $\lambda_{\min}(T^{-1}\sum_{t=1}^{T}\hat{G}_t\hat{G}_t') \geq c > 0$ with probability approach 1, which has been shown in the proof of Theorem A.1. Moreover, we have

$$\left\| T_j^{-1}\sum_{t=j+1}^{T}\varepsilon_t(\hat{G}_t - G_t)\right\|^2 \leq T_j^{-1}\sum_{t=1}^{T}\|\hat{G}_t - G_t\|^2\left(T_j^{-1}\sum_{t=j+1}^{T}\varepsilon_t^2\right),$$

where $\sum_{t=1}^{T}\|\hat{G}_t - G_t\|^2 \leq 2\sum_{t=1}^{T}\|\hat{G}_t - G_t(\hat{\theta})\|^2 + 2\sum_{t=1}^{T}\|G_t(\hat{\theta}) - G_t\|^2 = O_P(1)$, by Assumptions A.2–A.4 and the mean value theorem for the second term. It follows from Markov's inequality that

$$\begin{aligned} &T^{-1}\sum_{j=1}^{T-1}k^2(j/p)T_j\int\|\hat{B}_{4j}(v)\|^2\,dW(v) \\ &\quad\leq \max_{1\leq j<T}\sup_{v\in\mathbb{R}}\|\hat{\beta}_j(v)\|\int dW(v)\left(\sum_{t=1}^{T}\|\hat{G}_t - G_t\|^2\right)T^{-1}\sum_{j=1}^{T-1}k^2(j/p)T_j^{-1}\sum_{t=j+1}^{T}\varepsilon_t^2 \\ &\quad= O_P(p/T). \end{aligned} \quad \textbf{(A.44)}$$

Finally, we consider the last term $\hat{B}_{5j}(v)$ in (A.39). By the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \|\hat{B}_{5j}(v)\|^2 &\leq \left\|T_j^{-1}\sum_{t=j+1}^{T}G_t\varepsilon_t\right\|^2\|\hat{\beta}_j(v) - \beta_j(v)\|^2 \\ &\leq \left(T^{-1}\sum_{t=1}^{T}\|G_t\|^2\right)\left(T_j^{-1}\sum_{t=1}^{T}\varepsilon_t^2\right)(T/T_j)\|\hat{\beta}_j(v) - \beta_j(v)\|^2. \end{aligned}$$

We now focus on $\hat{\beta}_j(v) - \beta_j(v)$. Noting that $\beta_j(v) = [E(G_tG_t')]^{-1}\eta_j(v)$, we decompose

$$\begin{aligned} \hat{\beta}_j(v) - \beta_j(v) &= \left(T^{-1}\sum_{t=1}^{T}\hat{G}_t\hat{G}_t'\right)^{-1}T^{-1}\sum_{t=j+1}^{T}\hat{G}_t\hat{\psi}_{t-j}(v) - \beta_j(v) \\ &= \left\{\left(T^{-1}\sum_{t=1}^{T}\hat{G}_t\hat{G}_t'\right)^{-1} - [E(G_tG_t')]^{-1}\right\}T^{-1}\sum_{t=j+1}^{T}\hat{G}_t\hat{\psi}_{t-j}(v) \\ &\quad + [E(G_tG_t')]^{-1}\left[T^{-1}\sum_{t=j+1}^{T}\hat{G}_t\hat{\psi}_{t-j}(v) - \eta_j(v)\right] \\ &\equiv \hat{L}_{1j}(v) + \hat{L}_{2j}(v), \quad \text{say}. \end{aligned} \quad \textbf{(A.45)}$$

Using (A.8) and (A.9) and Chebyshev's inequality, we have $T^{-1}\sum_{t=1}^{T}\hat{G}_t\hat{G}_t' - E(G_tG_t') = O_P(T^{-1/2})$. This implies $(T^{-1}\sum_{t=1}^{T}\hat{G}_t\hat{G}_t')^{-1} - [E(G_tG_t')]^{-1} = O_P(T^{-1/2})$ given that $E(G_tG_t')$ is $O(1)$ and nonsingular. Moreover, given $|\hat{\psi}_{t-j}(v)| \leq 2$, we have $\max_{1\leq j\leq T}\sup_{v\in\mathbb{R}}\|T^{-1}\sum_{t=j+1}^{T}\hat{G}_t\hat{\psi}_{t-j}(v)\| \leq 2T^{-1}\sum_{t=1}^{T}\|\hat{G}_t\| = O_P(1)$. It follows that

$$\max_{1\leq j<T}\sup_{v\in\mathbb{R}}\|\hat{L}_{1j}(v)\| = O_P(T^{-1/2})O_P(1) = O_P(T^{-1/2}).$$

Therefore, we obtain

$$T^{-1}\sum_{j=1}^{T-1} k^2(j/p)T_j \int (T/T_j)\|\hat{L}_{1j}(v)\|^2\, dW(v) = O_P(p/T), \quad \textbf{(A.46)}$$

where we used the fact that $\sum_{j=1}^{T-1} k^2(j/p)T_j^{-1} = O(p/T)$.

Next, we consider $\hat{L}_{2j}(v)$, for which it suffices to consider $T^{-1}\sum_{t=j+1}^{T} \hat{G}_t\hat{\psi}_{t-j}(v) - \eta_j(v)$. We write

$$T^{-1}\sum_{t=j+1}^{T} \hat{G}_t\hat{\psi}_{t-j}(v) - \eta_j(v) = \frac{T_j}{T}\sum_{d=1}^{4} \hat{H}_{dj}(v) - \frac{j}{T}\eta_j(v),$$

where the $\hat{H}_{dj}(v)$ are defined as in (A.11). As shown in the proof of Theorem A.1, we have

$$\sum_{d=1}^{4} T^{-1}\sum_{j=1}^{T-1} k^2(j/p)T_j \int \|\hat{H}_{dj}(v)\|^2\, dW(v) = O_P(p/T).$$

Moreover, given $|k(\cdot)| \le 1$, we have

$$T^{-1}\sum_{j=1}^{T-1} k^2(j/p)T_j \int \left\| \frac{j}{T}\eta_j(v)\right\|^2 dW(v) \le T^{-2}\sum_{j=-\infty}^{\infty} j^2 \int \|\eta_j(v)\|^2\, dW(v) = O_P(T^{-2}),$$

where $\sum_{j=-\infty}^{\infty} j^2 \sup_{v\in\mathbb{R}}\|\eta_j(v)\|^2 < \infty$ given Assumption A.5 and the standard mixing inequality $\|\eta_j(v)\| \le C\alpha(j)^{(\nu-1)/\nu}$. It follows that

$$\begin{aligned} &T^{-1}\sum_{j=1}^{T-1} k^2(j/p)T_j(T/T_j) \int \|\hat{L}_{2j}(v)\|^2\, dW(v) \\ &\quad = O_P(p/T) + O_P(T^{-2}) = O_P(p/T). \end{aligned} \quad \textbf{(A.47)}$$

Combining (A.45)–(A.47), we obtain

$$T^{-1}\sum_{j=1}^{T-1} k^2(j/p)T_j \int \|\hat{B}_{5j}(v)\|^2\, dW(v) = O_P(p/T). \quad \textbf{(A.48)}$$

Combining (A.39)–(A.44) and (A.48) then yields

$$\hat{A}_3 \le T^{-1}2^4 \sum_{d=1}^{5}\sum_{j=1}^{T-1} k^2(j/p)T_j \int \|\hat{B}_{dj}(v)\|^2\, dW(v) = O_P(p/T) = o_P(1),$$

given $p/T \to 0$. This completes the proof of Theorem A.3. ■

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