Testing for the martingale difference hypothesis in multivariate time series models

Guochang Wang
College of Economics, Jinan University, Guangzhou, China

Ke Zhu
The University of Hong Kong

TESTING FOR THE MARTINGALE DIFFERENCE HYPOTHESIS IN MULTIVARIATE TIME SERIES MODELS

By Guochang Wang*, Ke Zhu† and Xiaofeng Shao‡

*Jinan University**, *University of Hong Kong*†
and University of Illinois at Urbana-Champaign‡

This paper proposes a general class of tests to examine whether the error term is a martingale difference sequence in a multivariate time series model with parametric conditional mean. These new tests are formed based on recently developed martingale difference divergence matrix (MDDM), and they provide formal tools to test the multivariate martingale hypothesis in the literature for the first time. Under suitable conditions, the asymptotic null distributions of these MDDM-based tests are established. Moreover, these MDDM-based tests are consistent to detect a broad class of fixed alternatives, and have nontrivial power against local alternatives of order $n^{-1/2}$, where n is the sample size. Since the asymptotic null distributions depend on the data generating process and the parameter estimation, a wild bootstrap procedure is further proposed to approximate the critical values of these MDDM-based tests, and its theoretical validity is justified. Finally, the usefulness of these MDDM-based tests is illustrated by simulation studies and one real data example.

1. Introduction. Consider a multivariate time series

$$Y_t = m(\mathcal{I}_{t-1}) + \varepsilon_t, \tag{1.1}$$

where $Y_t \in \mathcal{R}^p$, $m(\mathcal{I}_{t-1}) = E(Y_t|\mathcal{I}_{t-1})$ is the conditional mean almost surely (a.s.) of Y_t given the conditioning set $\mathcal{I}_{t-1}$, $\varepsilon_t = Y_t - E(Y_t|\mathcal{I}_{t-1})$ by construction is a martingale difference sequence (MDS) with respect to $\mathcal{I}_{t-1}$, and the conditioning set at time t is $\mathcal{I}_t = \{Y_t, Y_{t-1}, ...\}$. In parametric multivariate time series modeling, the form of $m(\cdot)$ is usually specified up to a finite dimensional parameter, including the vector auto-regressive moving average (VARMA) model in Lütkepohl (2005) and other nonlinear conditional mean models in Tsay (1998), Teräsvirta et al. (2010), Teräsvirta and Yang (2014), and

Keywords and phrases: martingale difference divergence matrix; martingale difference hypothesis; multivariate time series models; specification test; wild bootstrap

Dahlhaus (2017), to name a few. These linear/nonlinear multivariate conditional mean models have been widely applied in empirical macroeconomics and statistics.

In general, the empirical researchers assume that $m(\cdot) \in \mathcal{M}$, where $\mathcal{M} = \{f(\cdot, \theta) : \theta \in \Theta\}$ is a family of real functions indexed by an unknown vector parameter θ which lies in the parameter space $\Theta \in \mathcal{R}^s$. Since the inference of θ and the prediction of future values of Y_t heavily rely on the form of $m(\cdot)$, an important question is whether the form of $m(\cdot)$ is correctly specified. To answer this question, we need to test the hypothesis of $m(\cdot) \in \mathcal{M}$, that is,

$$H_0 : E(Y_t|\mathcal{I}_{t-1}) = f(\mathcal{I}_{t-1}, \theta_0) \text{ a.s. for some } \theta_0 \in \Theta,$$

against the alternative hypothesis that

$$H_1 : P\left(E(Y_t|\mathcal{I}_{t-1}) = f(\mathcal{I}_{t-1}, \theta)\right) < 1 \text{ for all } \theta \in \Theta.$$

Equivalently, the null hypothesis H_0 can be expressed as a martingale difference hypothesis (MDH):

$$H_0 : E(e_t|\mathcal{I}_{t-1}) = 0 \text{ a.s.} \tag{1.2}$$

(i.e., e_t is an MDS with respect to $\mathcal{I}_{t-1}$), where $e_t = e_t(\theta_0)$ with $e_t(\theta) = Y_t - f(\mathcal{I}_{t-1}, \theta)$. The concept of MDS plays a central role in many fundamental economic theories such as the efficient market hypothesis, rational expectations, or optimal consumption smoothing (see Hall (1978) and Lo (1997)). If H_0 is true, the best predictor, in the sense of least mean squared error, of the future values of Y_t given the conditioning set $\mathcal{I}_{t-1}$ is $f(\mathcal{I}_{t-1}, \theta_0)$. If H_0 is not true, there is a lack of fit in the postulated conditional mean specification $f(\mathcal{I}_{t-1}, \theta_0)$, which can lead to misleading statistical inferences and suboptimal point forecasts, resulting in erroneous conclusions. Hence, it is important to provide valid testing tools for H_0.

Since Durlauf (1991), testing for MDH has drawn much attention in the literature for the univariate time series Y_t (i.e., $p = 1$). The earlier proposals mainly considered the case that

$$f(\mathcal{I}_{t-1}, \theta_0) \equiv \mu \text{ (a constant scalar) a.s.} \tag{1.3}$$

Under H_0 and (1.3), past information does not help to improve the forecast of future values of Y_t. In this case, the spectral tests, which target nonzero serial correlations, were investigated by Durlauf (1991), Hong (1996), Deo (2000), and Shao (2011a), among others. However, these spectral tests are tailored to testing the lack of serial correlation but not for the MDH, and they are neither able to detect the pairwise non-MDS e_t (i.e., $E(e_t|Y_{t-i}) = 0$ for any $i \geq 1$ but $E(e_t|Y_{t-i}, Y_{t-j}) \neq 0$ for some $i \neq j$), nor consistent against the non-MDS e_t with zero autocorrelations. To overcome the first deficiency, Domínguez and Lobato (2003) and Kuan and Lee (2004) proposed two different classes of tests to detect the non-MDS e_t with respect to $\{Y_{t-1}, ..., Y_{t-M}\}$ for some integer $M \geq 1$. To overcome the second deficiency, Hong (1999) and Escanciano and Velasco (2006) constructed the generalized spectral test by using the characteristic functions; see also Bierens and Ploberger (1997) and Stinchcombe and White (1998) for related works. The generalized spectral test can detect the non-MDS e_t consistently but not the pairwise non-MDS e_t.

Since a constant conditional mean of Y_t in (1.3) may be very restrictive in many applications, more spectral tests for MDH are proposed in the presence of conditional mean dynamics, that is,

$$f(\mathcal{I}_{t-1}, \theta_0) \not\equiv \mu \ \text{ a.s.} \tag{1.4}$$

In this situation, the unknown parameter value θ_0 has to be replaced by its estimator. As a result, the spectral tests may need to take this estimation effect into account; see, for example, Hong and Lee (2005) for the kernel-based spectral test and Escanciano (2006) for the generalized spectral test. Indeed, the estimation effect has no impact on the limiting null distribution of the kernel-based spectral test in Hong and Lee (2005), while it does impact on the limiting null distribution of the generalized spectral test in Escanciano (2006). For other spectral tests on serial dependence hypotheses or model checking, we refer to Hong (1996), Paparoditis (2000), Delgado et al. (2005), Shao (2011b), Chen and Hong (2014), Zhu and Li (2015) and many others.

Although all of aforementioned tests have been shown useful in applications, they are only designed for the univariate MDH in (1.2). In practice, a univariate specification is inadequate to describe the dynamic behavior of multiple economic and financial variables with a causal relationship. This motivates the researchers to propose valid tests for the multivariate MDH (i.e., $p > 1$). Since the MDH implies the uncorrelatedness of e_t, one can find certain evidence to reject the multivariate MDH by using the residual-autocorrelations-based tests, see, for example, the portmanteau tests in Hosking (1980) and Li and McLeod (1981) for a stationary VARMA model. However, these residual-autocorrelation-based tests are expected to exhibit a lack of power to detect the non-MDS e_t with zero autocorrelations, and their generalizations to other nonlinear multivariate models are not straightforward. It is therefore important to explore general tests for the multivariate MDH in (1.2), yet no formal attempts have been made towards this goal.

This paper fills this gap by proposing some new one-sided tests for the multivariate MDH in (1.2) based on the martingale difference divergence matrix (MDDM) in Lee and Shao (2018). Our MDDM-based tests examine whether e_t is an MDS with respect to the user-chosen variable $K_t \in \mathcal{I}_{t-1}$. The choice of K_t is flexible and it can include the lagged dependent variables Y_{t-m} (for $m = 1, ..., M$) and their functional forms. Under suitable conditions, the asymptotic null distributions of our MDDM-based tests are established. Since the limiting null distributions are not pivotal, we follow Escanciano (2006) to implement a multivariate fixed-design wild bootstrap (FDWB) method to obtain the critical values of our MDDM-based tests, and provide a rigorous justification for its validity. Furthermore, it is shown that our MDDM-based tests are able to detect fixed alternatives consistently and possess nontrivial power against local alternatives of order $n^{-1/2}$, where n is the sample size. Our testing methodologies are applicable to the general specifications of $f(\cdot, \theta)$. By choosing different K_t, our MDDM-based tests can give investigators the complete information on the conditional mean structure of Y_t. Finally, the finite sample size and power performance of our MDDM-based tests is illustrated by simulation studies and a real data example.

The remaining paper is organized as follows. Section 2 introduces our MDDM-based test statistics. Section 3 gives the technical assumptions and studies the asymptotic properties of our MDDM-based tests. A wild bootstrap method is provided in Section 4. Simulation results are reported in Section 5. A real data example is presented in Section 6. Concluding remarks are offered in Section 7. The proofs are gathered in the Appendix, which is given in the Supplementary materials.

Throughout the paper, $\mathcal{R}$ is one-dimensional real vector space, $\mathcal{C}$ is one-dimensional complex vector space, $x^\star$ is used for "x-conjugate-transpose" (conjugate for scalars), $\|x\|^2 = x_1^2 + \ldots + x_p^2$ for $x \in \mathcal{R}^p$, $|x|_p^2 = x_1 x_1^\star + \ldots + x_p x_p^\star$ for $x \in \mathcal{C}^p$, and $\langle x, y\rangle$ is the inner product for $x, y \in \mathcal{R}^p$. For a matrix $X \in \mathcal{R}^{p\times q}$, X^T is its transpose, $tr(X)$ is its trace when $p = q$, $vec(X)$ is its vectorization, $\|X\|_F$ is its Frobenius norm, and $\|X\|_2$

is its spectral norm. Also, C is a generic constant, I_p is the $p \times p$ identity matrix, $I(\cdot)$ is the indicator function, n is the sample size, all limits are taken as $n \to \infty$, $o_p(1)(O_p(1))$ denotes a sequence of random vectors converging to zero (bounded) in probability, "$\to_p$" denotes the convergence in probability, and "$\to_d$" denotes the convergence in distribution.

2. The MDDM-based test statistics. In this section, we first introduce the martingale difference divergence matrix (MDDM) in Lee and Shao (2018), which forms the basis of our test statistics for H_0 in (1.2). For $V \in \mathcal{R}^p$ and $U \in \mathcal{R}^q$, the MDDM is defined as follows:

$$\mathrm{MDDM}(V|U) = \frac{1}{c_q} \int_{\mathcal{R}^q} \frac{\mathcal{G}(s)\mathcal{G}(s)^\star}{\|s\|^{1+q}} ds,$$

where $\mathcal{G}(s) = \mathrm{cov}(V, e^{i\langle s,U\rangle}) = (\mathcal{G}_1(s), ..., \mathcal{G}_p(s))^T$ for $s \in \mathcal{R}^q$, $\mathcal{G}_j(s) = \mathrm{cov}(V_j, e^{i\langle s,U\rangle})$, and $c_q = \pi^{(1+q)/2}/\Gamma((1+q)/2)$. If $E[\|U\|^2 + \|V\|^2] < \infty$, Lemma 3.1 in Lee and Shao (2018) asserts that $\mathrm{MDDM}(V|U)$ is a real, symmetric, and positive semidefinite matrix in $\mathcal{R}^{p\times p}$, and it has a simple and equivalent expression given by

$$\mathrm{MDDM}(V|U) = -E[(V - E(V))(V' - E(V'))^T \|U - U'\|], \tag{2.1}$$

where (V', U') is an independent and identically distributed (i.i.d.) copy of (V, U). For the sake of completeness, the proof of this equivalence is provided in the Appendix.

The result (2.1) implies that the sample MDDM (denoted by $\mathrm{MDDM}_n(V|U)$), based on an effective random sample $\{(U_k, V_k)\}_{k=n_0}^n$ from the joint distribution of (U, V), can be computed as

$$\mathrm{MDDM}_n(V|U) = -\frac{1}{N^2} \sum_{h,l=n_0}^{n} (V_h - \bar{V}_n)(V_l - \bar{V}_n)^T \|U_h - U_l\|,$$

where $n_0 \geq 1$ is a fixed integer, $N = n - n_0 + 1$ is the effective sample size, and $\bar{V}_n = N^{-1} \sum_{i=n_0}^n V_i$ (see, e.g., Lee and Shao (2018)).

The MDDM is a matrix-valued extension of the MDD^2 in Shao and Zhang (2014), and it has the following appealing property:

$$\mathrm{MDDM}(V|U) = 0 \ \text{ if and only if } \ E(V|U) = E(V) \text{ a.s.}$$

Hence, we can test the hypothesis of $E(V|U) = 0$ by examining whether $\|\mathrm{MDDM}_n(V|U)\|_F$ is significantly different from zero. In view of this fact, we consider the test statistic $\widehat{T}_n^F$

for H_0, where

$$\widehat{T}_n^F = N\|\mathrm{MDDM}_n(\widehat{e}_t|K_t)\|_F \tag{2.2}$$

for a k-dimensional random vector $K_t \in \mathcal{I}_{t-1}$, and k is fixed throughout the paper. Here,

$$\widehat{e}_t = Y_t - f(\widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n) \tag{2.3}$$

is the residual, $\widehat{\mathcal{I}}_t$ is the observed conditioning set at time t, and $\widehat{\theta}_n$ is an estimator of θ_0. To avoid any confusion, it pays to elaborate on the form of $\widehat{\mathcal{I}}_t$, which does not depend on any estimated quantity and is used to cope with the effect of unobserved initial values $\{Y_t\}_{t\le 0}$. Indeed, $\widehat{\mathcal{I}}_t$ is defined just as $\mathcal{I}_t$ but with the unobserved initial values $\{Y_t\}_{t\le 0}$ replaced by some given constants (e.g., $\{Y_t\}_{t\le 0} \equiv 0$). For example, we consider an invertible VARMA$(1,1)$ model: $Y_t = A_0 + A_1 Y_{t-1} + A_2\varepsilon_{t-1} + \varepsilon_t$, where $A_2 \not\equiv 0$. For this model, $f(\mathcal{I}_{t-1}, \theta) = A_0 + A_1 Y_{t-1} + A_2\varepsilon_{t-1}$, and it can be iteratively computed by

$$\begin{aligned} f(\mathcal{I}_{t-1}, \theta) = &\sum_{j=1}^{t} (-1)^{j-1} A_2^{j-1} A_0 + \sum_{j=1}^{t-1} (-1)^{j-1} A_2^{j-1}(A_1 + A_2) Y_{t-j} \\ &+ (-1)^{t-1} A_2^{t-1} A_1 Y_0 + (-1)^{t-1} A_2^{t} \varepsilon_0, \end{aligned}$$

where Y_0 and ε_0 depend on all unobserved initial values $\{Y_t\}_{t\le 0}$. To compute $f(\mathcal{I}_{t-1}, \theta)$, we can set all initial values $\{Y_t\}_{t\le 0} \equiv 0$, and consequently, $f(\mathcal{I}_{t-1}, \theta)$ is approximately computed by $f(\widehat{\mathcal{I}}_{t-1}, \theta)$, where

$$f(\widehat{\mathcal{I}}_{t-1}, \theta) = \sum_{j=1}^{t} (-1)^{j-1} A_2^{j-1} A_0 + \sum_{j=1}^{t-1} (-1)^{j-1} A_2^{j-1}(A_1 + A_2) Y_{t-j}$$

with $\widehat{\mathcal{I}}_{t-1} = \{Y_{t-1}, ..., Y_1, 0, 0, ...\}$. By doing this, the values of elements in $\widehat{\mathcal{I}}_t$ are all observed or known for computational purpose, and we thus call $\widehat{\mathcal{I}}_t$ the observed conditioning set. Note that the same notation has been used in Hong and Lee (2005, Assumption A5) and Escanciano (2006, Assumption A5).

Under H_0, $\mathrm{MDDM}(e_t|K_t) = 0$, so a large value of $\widehat{T}_n$ implies the rejection of H_0. Besides the Frobenius norm, we can also define our test statistic $\widehat{T}_n^S = N\|\mathrm{MDDM}_n(\widehat{e}_t|K_t)\|_2$ by using the spectral norm. Since the asymptotic properties of our test statistics based on both norms are similar, we shall only focus on the Frobenius norm for the presentation of our theory but we will examine the finite sample performance for both norms.

The choices of K_t in (2.2) are flexible for practitioners to study a complete conditional mean structure of Y_t. For example, we can choose $K_t = Y_{t,1:M} := (Y_{t-1}^T, ..., Y_{t-M}^T)^T$, which is the stacked vector of the lagged dependent variables Y_{t-m} for $m = 1, ..., M$, and the lag $M \geq 1$ is a fixed integer. This leads to an MDDM-based test statistic

$$\widehat{T}^F_{sn}(M) = N\|\mathrm{MDDM}_n(\widehat{e}_t|Y_{t,1:M})\|_F. \tag{2.4}$$

The idea of setting $K_t = Y_{t,1:M}$ comes from Domínguez and Lobato (2003) and Kuan and Lee (2004), and it aims to detect the general non-MDS structure of Y_t. Alternatively, we can adopt the pairwise approach by taking the weighted sum of $\|\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})\|_F$ for $j = 1, ..., M$. Using the weight sequence corresponding to the generalized spectral test in Escanciano (2006), we can form another MDDM-based test statistic

$$\widehat{T}^F_{wn}(M) = N\sum_{j=1}^{M}\omega_j\|\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})\|_F, \tag{2.5}$$

where $\omega_j = n_j/(Nj^2)$ and $n_j = N - j + 1$. Of course, all the Frobenius norm above can be replaced by the spectral norm. Thus, we have four test statistics, $\widehat{T}^F_{sn}$, $\widehat{T}^F_{wn}$, $\widehat{T}^S_{sn}$, and $\widehat{T}^S_{wn}$.

We emphasize that the selection of K_t shall meet user-specific needs. Besides the choices in (2.4)-(2.5), many other forms of K_t can also be used in practice. For instance, if the time series exhibits strong seasonality, it might be of interest to include some seasonal lagged variables into K_t, say, $K_t = (Y^T_{t-1}, ..., Y^T_{t-M}, Y^T_{t-d}, Y^T_{t-2d}, ..., Y^T_{t-Gd})^T$, where d is the known period, $M < d$, and G is the number of seasonal periods we include. We also note that the proposed MDDM-based tests bear some similarity to the Cramér-von Mises (CvM) type test developed in Escanciano (2006) in terms of the model checking framework, the methodology, technical assumptions, and related theories. However, there are some substantial differences to be elaborated below.

Escanciano's test is founded on the idea of generalized spectral test first formulated in Hong (1999), who adopted a generalized spectral density based approach and used a lag-window type spectral density estimator, which involves a kernel function and a smoothing parameter (or truncation lag). In Hong's test, the estimation effect (i.e., replacing θ_0 by $\widehat{\theta}_n$) is asymptotically negligible and the limiting null distribution is standard normal; see, for example, Hong and Lee (2005) for details. Escanciano's test is based on generalized spectral distribution function and involves no smoothing parameter, which might be convenient in practice, considering the difficulty in selecting the smoothing parameter for practitioners. However, the estimation effect is no longer asymptotically negligible and the limiting null distribution is nonpivotal, so typically a wild bootstrap approximation is used to approximate the critical values. Theoretically, the latter test has nontrivial power against local alternatives of order $n^{-1/2}$, which Hong's smoothing test cannot detect. Note that both Hong's test and Escanciano's test capture conditional mean dependence of e_t on some $Y_{t-j} \in \mathcal{I}_{t-1}$ for $j = 1, 2, ...$, and they can be viewed as a pairwise approach.

Our test statistic $\widehat{T}_{sn}^F$ captures the conditional mean dependence of e_t on $Y_{t,1:M}$ so can be regarded as a joint approach, whereas the variant $\widehat{T}_{wn}^F$ takes the pairwise perspective, and is in a sense like portmanteau test (Box and Pierce (1970)), which was briefly mentioned in the final section of Escanciano (2006). Note that Escanciano's test statistic has the following form

$$D_{n,W}^2(\widehat{\theta}_n) = \sum_{j=1}^{n} \frac{n_j}{(j\pi)^2} \int_\Gamma |\widehat{\gamma}_{j,w}(x, \widehat{\theta}_n)|^2 W(dx),$$

where $W(\cdot)$ is an absolutely continuous cumulative distribution function (cdf) with respect to Lebesgue measure, and $\widehat{\gamma}_{j,w}(x, \widehat{\theta}_n)$ is the sample version of $\gamma_{j,w}(x, \theta_0) := E(e_t w(Y_{t-j}, x))$. Since $\int_\Gamma |\gamma_{j,w}(x, \theta_0)|^2 W(dx) = 0$ if and only if $E(e_t|Y_{t-j}) = 0$, $\int_\Gamma |\gamma_{j,w}(x, \theta_0)|^2 W(dx)$ is a proper metric (analogous to MDD^2) of the conditional mean dependence of e_t upon Y_{t-j}, so in this sense the test statistic in Escanciano basically corresponds to a weighted sum of his conditional mean dependence metric at lag j.

In the univariate case $V \in \mathcal{R}$, $\text{MDDM}(V|U) = \text{MDD}^2(V|U)$. Hence, it is not difficult to see that our test statistic $\widehat{T}_{wn}^F$ differs from $D_{n,W}^2(\widehat{\theta}_n)$ in two aspects. On one hand, we use a fixed truncation lag M, whereas he used all the lags but the weights (over lags) he used discount higher-order lags severely so with some proper choice of M, the difference in terms of weights over lags is not large; see Escanciano (2009) for an in-depth discussion of the behavior of CvM test and the impact of M. On the other hand, the conditional mean dependence metric we used (i.e., the MDD) differs from his in somewhat fundamental ways. Our MDD has its roots from distance covariance (Szekely et al. (2007)), and uses the weighting function $(c_q \|s\|^{1+q})^{-1}$ in measuring the distance of $E(Ve^{i<s,U>})$ from $E(V)E(e^{i<s,U>})$ (reduced to 0 if $E(V) = 0$). In Escanciano's formulation, $w(Z, x)$ can take several different forms with $w(Z, x) = e^{ix^T Z}$ and $w(Z, x) = I(Z \le x)$ being two commonly used ones. If $w(Z, x) = e^{ix^T Z}$ is used, then the only difference from our MDD is the weighting function used. Our MDD uses a nonintegrable weighting function, whereas Escanciano (2006) assumed W to be a proper probability distribution function (e.g., the cdf of standard normal). As mentioned in Székely et al. (2007), one advantage of the nonintegrable weighting function used for distance covariance is that it yields scale invariance of distance correlation, which implies that if we multiply Y_t by a scalar $c > 0$, our test is invariant. It is not difficult to see that this invariance is lost for Escanciano's test if we let $w(Z, x) = e^{ix^T Z}$ and W be the cdf of standard normal distribution (see the supporting numerical evidences in the Supplementary materials). It is worth noting that if we let $w(Z, x) = I(Z \leq x)$, then the invariance is preserved even though the integrable weighting function W is used. However, the indicator function has the drawback of being more sensitive to the dimension than the characteristic function (see Escanciano (2006)). Another (less critical) difference is that the test in Escanciano (2006) is developed for the univariate Y_t, although an extension to the multivariate case seems possible but is not developed. Our test is directly formulated in a multivariate setting based on the MDDM.

The difference in the form of test statistics leads to different theoretical treatments. In Escanciano (2006), the key is to derive the weak convergence of the two-parameter process $S_{n,w}(\lambda, x, \widehat{\theta}_n) = \sum_{j=1}^{n} n_j^{1/2} \widehat{\gamma}_{j,w}(x, \widehat{\theta}_n) \frac{\sqrt{2}\sin(j\pi\lambda)}{j\pi}$ which lies in the Hilbert space $L_2(\Pi, v)$ of all complex-valued and squared v-integrable functions on $\Pi = [0, 1] \times \Gamma$, where v is the product measure of the W-measure and the Lebesgue measure on $[0, 1]$. By contrast, such arguments seem not directly transferrable to our setting, due to the non-integrability of our weighting function. Instead, we need to adopt a truncation argument and show the weak convergence of the process $\sqrt{n}\widehat{\mathcal{G}}_n(s)$ on any compact set Ω, where $\widehat{\mathcal{G}}_n(s)$ is the sample version of $\mathcal{G}(s)$ [see the beginning of Section 2 for the definition] applied to $\{\widehat{e}_t, K_t\}_{t=n_0}^{n}$ and is formally defined in (3.3) below. In particular, our tightness proof involves verifying the classical conditions in Bickel and Wichura (1971) for the weak convergence of multi-parameter stochastic process, and it seems quite different from the tightness proof in $L_2(\Pi, v)$; see Theorem 1 in Escanciano and Velasco (2006) and its proof.

The testing problem we address can also be posed as a special case of conditional moment restriction (CMR) testing problem, which has a large literature in econometrics and statistics. In the CMR testing framework, the null hypothesis of interest is

$$H_0^{CM} : E(\rho(Z_t, \theta_0)|X_t) = 0 \ \text{ a.s. for some } \theta_0 \in \Theta,$$

where $X_t \in \mathcal{R}^{l_x}$ and $Z_t \in \mathcal{R}^{l_z}$ are observed data with fixed dimensions $l_x, l_z \geq 1$, and $\rho(\cdot, \cdot) : \mathcal{R}^{l_z} \times \Theta \to \mathcal{R}$ is a known function. When $\{(X_t^T, Z_t^T)^T\}_{t=1}^{n}$ are i.i.d. and $l_z = 1$, Härdle and Mammen (1993) formed a test for H_0^{CM} by measuring the distance between the fitted parametric mean under the null and the estimated nonparametric mean function under the alternative; Stute (1997) and Stute et al. (1998) used the marked empirical processes to construct tests for H_0^{CM}, which involve no nonparametric estimation, but have an asymptotically non-negligible estimation effect; Whang (2000) proposed a test for H_0^{CM} by using the multivariate indicator function, and this testing method was further extended by Whang (2001) to the case of $l_z > 1$.

When $\{(X_t^T, Z_t^T)^T\}_{t=1}^n$ are dependent time series, Chen et al. (2003) constructed an empirical likelihood (EL)-based test for H_0^{CM} in the spirit of Härdle and Mammen (1993), and its adaptive type version was further studied in Chen and Gao (2007) for the case $l_z = 1$. When $l_z > 1$, very little work has been done in this context except for Tripathi and Kitamura (2003), who considered the EL-based tests for H_0^{CM} by using the nonparametric kernel weights. Our testing framework naturally falls into the CMR framework by letting $\rho(Z_t, \theta_0) = Y_t - f(\mathcal{I}_{t-1}, \theta_0)$ and $X_t = K_t$, so the EL-based tests in Tripathi and Kitamura (2003) seem applicable to our testing problem. However, since these EL-based test use l_x-dimensional kernel weights, they neither detect local alternatives of order $n^{-1/2}$ nor perform well empirically for $l_x > 3$ as often demonstrated in nonparametric smoothing literature. Along a different line, Koul and Stute (1999) extended Stute's (1997) method (for $l_x = 1$ and $l_z = 1$) by using the martingale transformation of Khmaladze (1981), such that the resulting test for H_0^{CM} is asymptotically pivotal; Khmaladze and Koul (2004) gave an asymptotically pivotal test for H_0^{CM} by using innovation martingale transform idea of Khmaladze (1993), but this test requires i.i.d. model errors.

3. Asymptotic theory. In this section, we study the asymptotics of $\widehat{T}_n^F$ in (2.2). Similar results can be shown for $\widehat{T}_{sn}^F$, $\widehat{T}_{wn}^F$, $\widehat{T}_{sn}^S$, and $\widehat{T}_{wn}^S$ with a slight modification, and the details are omitted.

3.1. *Technical assumptions.* Let $\{Y_t\}_{t=1}^n$ be a sequence of observations from model (1.1), $\mathcal{F}_t := \sigma(\mathcal{I}_t)$ be the sigma-field generated by $\mathcal{I}_t$, and $g_t(\theta) := g(\mathcal{I}_{t-1}, \theta) = \partial f(\mathcal{I}_{t-1}, \theta)/\partial\theta'$. Write $K_t = (K_{1t}, ..., K_{kt})'$. We need the following five assumptions to derive the asymptotic properties of $\widehat{T}_n^F$.

ASSUMPTION 3.1. *(i) $\{(Y_t, \varepsilon_t, K_t)\}$ is strictly stationary and ergodic; (ii) $E\|\varepsilon_t\|^4 < \infty$; (iii) $E\|K_t\|^4 < \infty$; (iv) $E\prod_{l=1}^k \|K_{lt}\|^{2u} < \infty$ for some $u > 1$.*

ASSUMPTION 3.2. *The function $f(\mathcal{I}_{t-1}, \cdot)$ is twice continuously differentiable on Θ.*

The score $g_t(\theta)$ *satisfies* $E\sup_{\theta\in\Theta}\|g_t(\theta)\|_F^4<\infty$.

Assumption 3.3. *The parametric space* Θ *is compact in* $\mathcal{R}^s$. *The true parameter* θ_0 *is an interior point of* Θ. *There exists a unique* $\theta_*\in\Theta$ *such that* $\|\widehat{\theta}_n-\theta_*\|=o_p(1)$ *under both hypotheses* H_0 *and* H_1, *and* θ_* *can take different values under* H_0 *and* H_1.

Assumption 3.4. *The estimator* $\widehat{\theta}_n$ *given in (2.3) satisfies the asymptotic expansion under* H_0,

$$\sqrt{n}(\widehat{\theta}_n-\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\pi(Y_t,\mathcal{I}_{t-1},\theta_0)+o_p(1), \tag{3.1}$$

where $\pi(\cdot,\cdot,\theta)$ *is a measurable function in* $\mathcal{R}^s$, $E[\pi(Y_t,\mathcal{I}_{t-1},\theta_0)|\mathcal{F}_{t-1}]=0$, *and* $L(\theta_0)=E[\pi(Y_t,\mathcal{I}_{t-1},\theta_0)\pi(Y_t,\mathcal{I}_{t-1},\theta_0)^T]$ *exists and is positive definite.*

Assumption 3.5. *For* $\widehat{\mathcal{I}}_t$ *given in (2.3),*

$$\lim_{n\to\infty}\sum_{t=1}^{n}\left(E\sup_{\theta\in\Theta}\|f(\mathcal{I}_{t-1},\theta)-f(\widehat{\mathcal{I}}_{t-1},\theta)\|^4\right)^{1/4}<\infty.$$

Assumptions 3.1–3.5 are similar to those made in Escanciano (2006). Assumption 3.1 gives some technical conditions for the data generating process (DGP). Assumption 3.1(i) is quite standard in time series analysis, Assumption 3.1(ii) and (iv) are sufficient to prove the weak convergence of $\widehat{\mathcal{G}}_n(s)$ in Theorem 3.1 below, and Assumption 3.1(iii) is used to show the negligibility of the initial values. Our simulation results in the Supplementary materials suggest that the moment conditions in Assumption 3.1(ii)–(iv) are sufficient but not necessary for our asymptotic theory. Like Escanciano (2006), no mixing or asymptotic independence assumption is needed to derive our asymptotic theory. Assumption 3.2 requires some regularity conditions on the model, and it is standard in the conditional mean specification literature; see Bierens and Ploberger (1997) and Lütkepohl (2005). Assumption 3.3 requires the consistency of $\widehat{\theta}_n$, and under H_0, $\theta_*=\theta_0$. Under H_1, θ_* is not necessarily the same as θ_0, and it can be obtained as the parameter that minimizes certain distance between the true DGP and the misspecified parametric model; see White (1982). Assumption 3.4 is satisfied under mild conditions for most estimators; see, for example, Hall and Heyde (1980), Tjøstheim (1986), and Lütkepohl (2005). Assumption 3.5 is a condition on the approximation error by replacing the information set $\mathcal{I}_{t-1}$ by $\widehat{\mathcal{I}}_{t-1}$, and it guarantees that the unobserved initial values have the negligible effect on our asymptotic theory. Similar assumptions have been used in Escanciano (2006) and Hong and Lee (2005). In particular, Hong and Lee (2005) showed that Assumption 3.5 can be verified for an ARMA(1, 1) model.

3.2. *Asymptotic null distribution.* In this subsection, we derive the asymptotic distribution of $\widehat{T}_n^F$ under H_0. Following Shao and Zhang (2014), we re-write $\widehat{T}_n^F$ as

$$\widehat{T}_n^F = N \left\| \frac{1}{c_k} \int_{\mathcal{R}^k} \frac{\widehat{\mathcal{G}}_n(s)\widehat{\mathcal{G}}_n(s)^\star}{\|s\|^{1+k}} ds \right\|_F, \tag{3.2}$$

where

$$\widehat{\mathcal{G}}_n(s) = \frac{1}{N}\sum_{t=n_0}^{n} \widehat{e}_t e^{i\langle s,K_t\rangle} - \left(\frac{1}{N}\sum_{t=n_0}^{n} \widehat{e}_t\right)\left(\frac{1}{N}\sum_{t=n_0}^{n} e^{i\langle s,K_t\rangle}\right). \tag{3.3}$$

To elaborate the asymptotic distribution of $\widehat{T}_n^F$, we need more notation. Let $\Upsilon(s) = \mathrm{cov}(g_t(\theta_0), e^{i\langle s,K_t\rangle})$, $\mathcal{V}$ be a normal random vector with mean zero and variance-covariance matrix $L(\theta_0)$, $\Delta(s)$ be a complex valued Gaussian field with mean zero and covariance matrix function $\mathrm{cov}(\Delta(s), \Delta(s')^\star) = E(\varepsilon_t\varepsilon_t^T e^{i[\langle s,K_t\rangle + \langle s',K_t\rangle]})$, and $(\Delta(s), \mathcal{V})$ be jointly Gaussian with covariance matrix $\mathrm{cov}(\Delta(s), \mathcal{V}) = E(\varepsilon_t \pi(Y_t, \mathcal{I}_{t-1}, \theta_0)^T e^{i\langle s,K_t\rangle})$.

Our first theorem gives the weak convergence result of $\{\widehat{\mathcal{G}}_n(s), s \in \Omega\}$, where $\Omega \subset \mathcal{R}^k$ is a compact set. Note that we are dealing with a sequence of continuous complex-valued random functions over Ω. We shall use "$\Rightarrow$" to denote weak convergence in $\mathcal{C}$, where $\mathcal{C}$ stands for the space of continuous complex-valued random function over Ω equipped with uniform topology. In particular, the process convergence is denoted by $\widehat{\mathcal{G}}_n(\cdot) \Rightarrow \mathcal{G}(\cdot)$ or $\widehat{\mathcal{G}}_n(s) \Rightarrow \mathcal{G}(s)$ on Ω, where $\mathcal{G}(s)$ is a process on Ω.

THEOREM 3.1. *Suppose that Assumptions 3.1–3.5 hold. Then, under H_0,*

$$\sqrt{N}\widehat{\mathcal{G}}_n(s) \Rightarrow \chi(s) := \Delta(s) - \Upsilon(s)\mathcal{V},$$

on any compact set $\Omega \subset \mathcal{R}^k$.

Based on Theorem 3.1, we can establish the asymptotic null distribution of $\widehat{T}_n^F$ in the following corollary.

COROLLARY 3.1. *Suppose that Assumptions 3.1–3.5 hold. Then, under H_0,*

$$\widehat{T}_n^F \to_d \left\| \frac{1}{c_k} \int_{\mathcal{R}^k} \frac{\chi(s)\chi(s)^\star}{\|s\|^{1+k}} ds \right\|_F,$$

where $\chi(s)$ is defined as in Theorem 3.1.

At the significance level α, we set the rejection region as

$$\widehat{T}_n^F > c_\alpha, \tag{3.4}$$

where the critical value c_α is the αth upper percentile of the asymptotic null distribution of $\widehat{T}_n^F$ in Corollary 3.1. Note that the asymptotic null distribution of $\widehat{T}_n^F$ depends on the process $\chi(s)$. The first part of $\chi(s)$ reflects the unknown dependence structure of $\{(\varepsilon_t, K_t)\}$, and the second part of $\chi(s)$ involves the estimation effect resulting from $\widehat{\theta}_n$. Consequently, the asymptotic null distribution of $\widehat{T}_n^F$ is not pivotal, and its critical value c_α needs to be approximated by a bootstrap method (see Section 4 below).

3.3. *Consistency and local alternatives.* In this subsection, we first study the asymptotic behavior of $\widehat{T}_n^F$ under the following global alternative (see Escanciano (2006)):

$$H_a : Y_t = f(\mathcal{I}_{t-1}, \theta_*) + a_t + \varepsilon_t,$$

where $\{a_t\}$ is strictly stationary, ergodic, and $\mathcal{F}_{t-1}$-measurable. Theorem 3.2 shows the asymptotic behavior of $\widehat{\mathcal{G}}_n(s)$ under H_a.

THEOREM 3.2. *Suppose that $E\|a_t\| < \infty$ and Assumptions 3.1(i)–(ii), 3.2–3.3 and 3.5 hold. Then, under H_a,*

$$\sup_{s\in\Omega} \left| \widehat{\mathcal{G}}_n(s) - \mathcal{A}(s) \right| \to_p 0,$$

for any compact set $\Omega \subset \mathcal{R}^k$, where $\mathcal{A}(s) = cov(a_t, e^{i\langle s, K_t\rangle})$.

Let Ξ denote the class of alternatives $\{a_t\}$ for which it holds under H_a, $\mathcal{A}(s) \neq 0$ with positive P-measure. Then, if $a_t \in \Xi$,

$$\frac{1}{c_k}\int_{\bar{\Omega}} \frac{\widehat{\mathcal{G}}_n(s)\widehat{\mathcal{G}}_n(s)^\star}{\|s\|^{1+k}} ds \to_p \frac{1}{c_k}\int_{\bar{\Omega}} \frac{\mathcal{A}(s)\mathcal{A}(s)^\star}{\|s\|^{1+k}} ds > 0 \text{ (elementwise)},$$

for some compact set $\bar{\Omega} \subset \mathcal{R}^k$, and consequently the test statistic $\widehat{T}_n^F$ will be consistent against H_a. We shall mention that Ξ can cover a broad class of alternatives with respect to different choices of K_t, although it does not contain all possible alternatives.

Second, we study the asymptotic behavior of $\widehat{T}_n^F$ under the following local alternative (see Bierens and Ploberger (1997) and Escanciano (2006)):

$$H_{a,n} : Y_t = f(\mathcal{I}_{t-1}, \theta_0) + \frac{a_t}{\sqrt{n}} + \varepsilon_t,$$

where $\{a_t\}$ is defined as in H_a. To proceed further, we need an additional assumption regarding the behavior of the estimator $\widehat{\theta}_n$ under $H_{a,n}$, which was also assumed in Escanciano (2006).

ASSUMPTION 3.6. *The estimator $\widehat{\theta}_n$ given in (2.3) satisfies the asymptotic expansion under $H_{a,n}$,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \xi_a + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \pi(Y_t, \mathcal{I}_{t-1}, \theta_0) + o_p(1),$$

where $\pi(Y_t, \mathcal{I}_{t-1}, \theta_0)$ is defined as in Assumption 3.4 and $\xi_a \in \mathcal{R}^s$.

THEOREM 3.3. *Suppose that $E\|a_t\| < \infty$ and Assumptions 3.1–3.2 and 3.5–3.6 hold. Then, under $H_{a,n}$,*

$$\sqrt{N}\widehat{\mathcal{G}}_n(s) \Rightarrow \chi_a(s) := \chi(s) + \mathcal{A}(s) - \Upsilon(s)\xi_a,$$

on any compact set $\Omega \subset \mathcal{R}^k$, where $\chi(s)$ and $\Upsilon(s)$ are defined as in Theorem 3.1, and $\mathcal{A}(s)$ is defined as in Theorem 3.2.

Based on Theorem 3.3, we can obtain the asymptotic distribution of $\widehat{T}_n^F$ under $H_{a,n}$.

COROLLARY 3.2. *Suppose that $E\|a_t\|^4 < \infty$ and Assumptions 3.1–3.2 and 3.5–3.6 hold. Then, under $H_{a,n}$,*

$$\widehat{T}_n^F \to_d \left\| \frac{1}{c_k} \int_{\mathcal{R}^k} \frac{\chi_a(s)\chi_a(s)^\star}{\|s\|^{1+k}} ds \right\|_F,$$

where $\chi_a(s)$ is defined as in Theorem 3.3.

The aforementioned corollary implies that $\widehat{T}_n^F$ has the non-trivial local power when $\mathcal{A}(s) - \Upsilon(s)\xi_a \neq 0$. Specifically, if $\xi_a = 0$, $\widehat{T}_n^F$ has the non-trivial power against all local alternatives in Ξ, whereas if $\xi_a \neq 0$, $\widehat{T}_n^F$ has nontrivial power against the local alternatives in Ξ not collinear to the score $g_t(\theta_0)$.

Note that the asymptotic local power function (ALPF) of $\widehat{T}_n^F$ under $H_{a,n}$ is

$$\text{ALPF} = P\Big(\left\| \frac{1}{c_k} \int_{\mathcal{R}^k} \frac{\chi_a(s)\chi_a(s)^\star}{\|s\|^{1+k}} ds \right\|_F > c_\alpha \Big).$$

Because the ALPF depends on not only the null specification but also the model under the alternative, the estimation method, and the variable K_t, it seems challenging to work out its explicit value for a given alternative. For $\widehat{T}_{sn}^{F}$, $\widehat{T}_{wn}^{F}$, $\widehat{T}_{sn}^{S}$, and $\widehat{T}_{wn}^{S}$, similar conclusions hold, and their ALPFs indeed depend on the lag M (through K_t). Consequently, it is hard to study how the choice of M impacts the value of ALPF from a theoretical angle. This makes it difficult to provide a theory-guided choice of M for $\widehat{T}_{sn}^{F}$, $\widehat{T}_{wn}^{F}$, $\widehat{T}_{sn}^{S}$, and $\widehat{T}_{wn}^{S}$. In practice, when there is prior information about possible lag at which the serial linear/nonlinear dependence exists, we should use that information to select M. For example, if we analyze quarterly time series and expect no dependence beyond lag 4, then choosing $M = 4$ is appropriate. If there is no prior information, using all of lags up to a large lag $\bar{M}$ (i.e., $M = 1, ..., \bar{M}$) may be preferred to guard against possible dependence at all considered lags. Note that the portmanteau test and its variants often use this method to select their lag variable (see Li (2003)), which plays a similar role as our lag variable M.

4. Bootstrap approximations. Since the asymptotic null distribution of $\widehat{T}_{n}^{F}$ in Theorem 3.1 is not pivotal, its critical values have to be approximated. Similar to Escanciano (2006), we apply a fixed-design wild bootstrap (WB) method to approximate the critical values of $\widehat{T}_{n}^{F}$. The WB method was initiated by Wu (1986) and Liu (1988), and has been widely used in the literature of model checking for time series models; see Gonçalves and Kilian (2004), Escanciano (2006), Shao (2010), Zhu (2016) and references therein. Our fixed-design WB bootstrap procedure is given as follows:

1. Estimate the original model and obtain the residuals $\{\widehat{e}_t\}_{t=1}^{n}$ according to (2.3).
2. Generate WB residuals $\{\widehat{e}_t^{*}\}_{t=1}^{n}$ by $\widehat{e}_t^{*} = \widehat{e}_t w_t^{*}$, with $\{w_t^{*}\}$ being a sequence of i.i.d. random variables with mean zero, unit variance, and bounded support and also independent of the sequence $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^{n}$.
3. Given $\widehat{\theta}_n$ and $\widehat{e}_t^{*}$, generate bootstrap data $\{Y_t^{*}\}_{t=1}^{n}$ by $Y_t^{*} = f(\widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n) + \widehat{e}_t^{*}$.
4. Compute $\widehat{\theta}_n^{*}$ from the data $\{Y_t^{*}\}_{t=1}^{n}$ in the same way as for $\widehat{\theta}_n$, and then calculate the corresponding bootstrap residuals $\{\widehat{e}_t^{**}\}_{t=1}^{n}$ by $\widehat{e}_t^{**} = Y_t^{*} - f(\widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n^{*})$.
5. Compute the bootstrap MDDM-based test statistic $\widehat{T}_{n}^{F*}$ in the same way as for $\widehat{T}_{n}^{F}$ with $\widehat{e}_t$ replaced by $\widehat{e}_t^{**}$.
6. Repeat steps 2–5 B times to obtain $\{\widehat{T}_{nb}^{F*}\}_{b=1}^{B}$, and denote the αth upper percentile of $\{\widehat{T}_{nb}^{F*}\}_{b=1}^{B}$ by c_α^{*}, which is considered as the approximated value of c_α in (3.4).

The distributions of $\{w_t^{*}\}$ are flexible in practice. For instance, we can follow Mammen

(1993) to choose $\{w_t^*\}$ as i.i.d. random variables with the distribution

$$P\left(w_t^* = \frac{1-\sqrt{5}}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} \quad \text{and} \quad P\left(w_t^* = \frac{1+\sqrt{5}}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}};$$

or we can follow Liu (1988) to choose i.i.d. $\{w_t^*\}$ having Rademacher distribution

$$P(w_t^* = 1) = \frac{1}{2} \quad \text{and} \quad P(w_t^* = -1) = \frac{1}{2}.$$

Similar to (3.2), we have

$$\widehat{T}_n^{F*} = N \left\| \frac{1}{c_k} \int_{\mathcal{R}^k} \frac{\widehat{\mathcal{G}}_n^*(s)\widehat{\mathcal{G}}_n^*(s)^\star}{\|s\|^{1+k}} ds \right\|_F,$$

where $\widehat{\mathcal{G}}_n^*(s)$ is defined in the same way as $\widehat{\mathcal{G}}_n(s)$ in (3.3) with $\widehat{e}_t$ replaced by $\widehat{e}_t^{**}$. In order to justify the validity of our fixed-design WB method, we need one additional assumption, which is similar to Assumption A7 in Escanciano (2006), and is used to show the weak convergence of $\widehat{\mathcal{G}}_n^*(s)$. Below, we let E^* denote the expectation conditional on $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$, and $o_p^*(1)$ denote a sequence of random variables converging to zero in probability conditional on $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$. We also define $L^*(\widehat{\theta}_n) = E^*[\pi(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)\pi(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)^T]$.

Assumption 4.1. *The estimator $\widehat{\theta}_n^*$ satisfies the asymptotic expansion*

$$\sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \pi(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n) + o_p^*(1),$$

where $\pi(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)$ satisfies

(i) $E^*[\pi(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)] = 0$ a.s.;

(ii) $L^*(\widehat{\theta}_n)$ exists and is positive definite (a.s.) with $L^*(\widehat{\theta}_n) = L(\theta_*) + o_p^*(1)$, where $L(\theta_*) = E[\pi(Y_t, \mathcal{I}_{t-1}, \theta_*)\pi(Y_t, \mathcal{I}_{t-1}, \theta_*)^T]$;

(iii) $\frac{1}{N}\sum_{t=n_0}^n E^*\left[\widehat{e}_t^* e^{i\langle s, K_t\rangle}\pi(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)\right] = E\left[e_t(\theta_*)e^{i\langle s, K_t\rangle}\pi(Y_t, \mathcal{I}_{t-1}, \theta_*)\right] + o_p^*(1)$ on any compact set $\Omega \subset \mathcal{R}^k$.

We first show the weak convergence of $\widehat{\mathcal{G}}_n^*(s)$ in the following theorem.

Theorem 4.1. *Suppose that Assumptions 3.1–3.3, 3.5 and 4.1 hold. Then, conditional on $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$,*

$$\sqrt{N}\widehat{\mathcal{G}}_n^*(s) \Rightarrow \chi_*(s) := \Delta_*(s) - \Upsilon_*(s)\mathcal{V}_* \text{ in probability},$$

on any compact set $\Omega \subset \mathcal{R}^k$, where $\Delta_*(s)$, $\Upsilon_*(s)$, and $\mathcal{V}_*$ are defined in the same way as $\Delta(s)$, $\Upsilon(s)$, and $\mathcal{V}$ in Theorem 3.1 with θ_0 replaced by θ_*.

Next, based on the preceding theorem, Corollary 4.1 guarantees that our bootstrap critical value c_α^* computed from steps 1–6 is valid under the null hypothesis H_0, any fixed alternative hypothesis H_a, or the local alternative hypothesis $H_{a,n}$.

COROLLARY 4.1. *Suppose that Assumptions 3.1–3.3, 3.5 and 4.1 hold. Then, conditional on* $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$,

$$\widehat{T}_n^{F*} \to_d \left\| \frac{1}{c_k} \int_{\mathcal{R}^k} \frac{\chi_*(s)\chi_*(s)^\star}{\|s\|^{1+k}} ds \right\|_F \quad \textit{in probability,}$$

where $\chi_*(s)$ *is defined as in Theorem 4.1.*

In particular, the above limit under H_0 is the same as the limiting null distribution of $\widehat{T}_n^F$, implying the asymptotic size accuracy. Also, since $c_\alpha^* = O_p^*(1)$ under H_a or $H_{a,n}$, it is not hard to see that for $a_t \in \Xi$,

$$\lim_{n\to\infty} P\left(\widehat{T}_n^F > c_\alpha^*\right) = 1 \quad \text{under } H_a,$$

$$\lim_{\|E(a_t)\|\to\infty} \lim_{n\to\infty} P\left(\widehat{T}_n^F > c_\alpha^*\right) = 1 \quad \text{under } H_{a,n},$$

meaning that $\widehat{T}_n^F$ with the bootstrapped critical value c_α^* can detect H_a consistently and has nontrivial power to detect $H_{a,n}$.

5. Simulation studies. In this section, we carry out simulation experiments to assess the performance of our MDDM-based tests $\widehat{T}_{sn}^F$, $\widehat{T}_{wn}^F$, $\widehat{T}_{sn}^S$, and $\widehat{T}_{wn}^S$ in finite samples. We present some direct comparison with the tests in Hong and Lee (2005) and Escanciano (2006) when $p = 1$ and some comparison with the tests in Box and Pierce (1970), Hosking (1980), Li and Mcleod (1981), and Lütkepohl (2005) when $p > 1$. The DGPs considered below are the same as those used in Hong and Lee (2005) when $p = 1$ and multivariate counterparts of those univariate DGPs in Hong and Lee (2005) and Escanciano (2006) when $p > 1$. For all simulations, we take the sample size $n = 200$ and 500, choose the lag $M = 3$, 6, and 9, and set the significance level $\alpha = 5\%$. For the wild bootstrap method of our MDDM-based tests, we use the Rademacher distribution for w_t^* and do bootstrap $B = 1000$ times.

5.1. *Simulations for $p = 1$.* In this subsection, we compare our MDDM-based tests with the tests $\widehat{M}_i(p)$ (for $i = 1, 2, 3$) in Hong and Lee (2005) and the tests $D_{n,C}^2$ and $D_{n,I}^2$ in Escanciano (2006).

Our null model is a univariate AR(1) model:

$$Y_t = A_0 + A_1 Y_{t-1} + \varepsilon_t, \tag{5.1}$$

where $\varepsilon_t = v_t^{1/2}\eta_t$ and $v_t = \phi_1 + \phi_2\varepsilon_{t-1}^2$. To examine the size performance of all tests, we generate 1000 replications of sample size n from the following two DGPs based on the above model (5.1):

$$\text{DGP } 1: \phi_1 = 1 \text{ and } \phi_2 = 0;$$

$$\text{DGP } 2: \phi_1 = 0.43 \text{ and } \phi_2 = 0.57,$$

where $A_0 = 0$, $A_1 = 0.5$, and η_t is a sequence of i.i.d. $N(0,1)$ random variables. To examine the power performance of all tests, we generate 1000 replications of sample size n from the following eight DGPs:

$$\text{DGP } 3: Y_t = 0.5Y_{t-1} + 0.6Y_{t-1}\varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP } 4: Y_t = 0.5Y_{t-1} - 0.6\varepsilon_{t-1}^2 + \varepsilon_t;$$

$$\text{DGP } 5: Y_t = 0.5Y_{t-1} + 10Y_{t-1}\exp(-Y_{t-1}^2) + \varepsilon_t;$$

$$\text{DGP } 6: Y_t = 0.5Y_{t-1}I(Y_{t-1} \le 0) - 0.5Y_{t-1}I(Y_{t-1} > 0) + \varepsilon_t;$$

$$\text{DGP } 7: Y_t = 1 - 0.5Y_{t-1} - (4 + 0.4Y_{t-1})/(1 + \exp(-Y_{t-1})) + \varepsilon_t;$$

$$\text{DGP } 8: Y_t = 0.5Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP } 9: Y_t = 0.5Y_{t-1} + \sum_{j=1}^{5} 0.5^j \varepsilon_{t-j}^2 + \varepsilon_t;$$

$$\text{DGP } 10: Y_t = I(Y_{t-6} > 0) - I(Y_{t-6} < 0) + \varepsilon_t,$$

where $\varepsilon_t = \eta_t$.

For each replication, we fit it by model (5.1) and obtain the model residual $\widehat{e}_t$, where $\widehat{e}_t = Y_t - \widehat{A}_{0n} - \widehat{A}_{1n}Y_{t-1}$, and $\widehat{\theta}_n = (\widehat{A}_{0n}, \widehat{A}_{1n})'$ is the least squares estimate. Based on $\{\widehat{e}_t\}$, we calculate our MDDM-based test statistics $\widehat{T}_{sn}^F(M)$ and $\widehat{T}_{wn}^F(M)$. Here, we do not need to consider $\widehat{T}_{wn}^S(M)$ and $\widehat{T}_{sn}^S(M)$, which are identical to $\widehat{T}_{wn}^F(M)$ and $\widehat{T}_{sn}^F(M)$, respectively, in the case of $p = 1$. To make a comparison, we also calculate the test statistics $\widehat{M_i}(p)$ (for $i = 1, 2, 3$) in Hong and Lee (2005) and the test statistics $D_{n,C}^2$ and $D_{n,I}^2$ in Escanciano (2006).

Specifically, the test statistics $\widehat{M}_i(p)$ (for $i = 1, 2, 3$) are computed as

$$\widehat{M}_i(p) = \frac{\sum_{j=1}^{n-1} K^2(j/p)\{\int_{\mathcal{R}} |\sum_{t=j+1}^{n} \widehat{e}_t \widehat{\psi}_{t-j}(x)|^2 W(dx)\}/(n-j)^2 - \widehat{C}_i(p)}{\sqrt{\widehat{D}_i(p)}}$$

with $\widehat{\psi}_t(x) = e^{ix\widehat{e}_t} - \widehat{\phi}(x)$ and $\widehat{\phi}(x) = n^{-1}\sum_{t=1}^{n} e^{ix\widehat{e}_t}$, where the centering factors $\widehat{C}_i(p)$ and the scaling factors $\widehat{D}_i(p)$ are given by

$$\widehat{C}_1(p) = \sum_{j=1}^{n-1} \frac{K^2(j/p)}{n-j} \sum_{t=j+1}^{n} \widehat{e}_t^2 \int_{\mathcal{R}} |\widehat{\psi}_{t-j}(x)|^2 W(dx),$$

$$\widehat{C}_2(p) = \widehat{C}_3(p) = \widehat{\sigma}_e^2 \sum_{j=1}^{n-1} K^2(j/p) \int_{\mathcal{R}} (1 - |\widehat{\phi}(x)|^2) W(dx),$$

$$\begin{aligned}\widehat{D}_1(p) = {} & 2\sum_{j,l=1}^{n-2} \frac{K^2(j/p)K^2(l/p)}{(n - \max\{j,l\})^2} \\ & \times \int_{\mathcal{R}^2} \Big| \sum_{t=\max\{j,l\}+1}^{n} \widehat{e}_t^2 \widehat{\psi}_{t-j}(x)\widehat{\psi}_{t-l}(y) \Big|^2 W(dx)W(dy),\end{aligned}$$

$$\begin{aligned}\widehat{D}_2(p) = {} & 2\widehat{\sigma}_e^4 \sum_{j,l=1}^{n-2} \frac{K^2(j/p)K^2(l/p)}{(n - |j-l|)^2} \\ & \times \int_{\mathcal{R}^2} \Big| \sum_{t=|j-l|+1}^{n} e^{ix\widehat{e}_t}[e^{iy\widehat{e}_{t-|j-l|}} - \widehat{\phi}(y)] \Big|^2 W(dx)W(dy),\end{aligned}$$

$$\widehat{D}_3(p) = 2\widehat{\sigma}_e^4 \sum_{j=1}^{n-2} \frac{K^4(j/p)}{n^2} \int_{\mathcal{R}^2} \Big| \sum_{t=1}^{n} e^{ix\widehat{e}_t}[e^{iy\widehat{e}_t} - \widehat{\phi}(y)] \Big|^2 W(dx)W(dy)$$

with $\widehat{\sigma}_e^2 = n^{-1}\sum_{t=1}^{n} \widehat{e}_t^2$. As shown in Hong and Lee (2005), $\widehat{M}_1(p)$, $\widehat{M}_2(p)$, and $\widehat{M}_3(p)$ work for conditional heteroscedastic, conditional homoscedastic, and i.i.d. errors, respectively, and all of them have the standard normal limiting null distribution. To compute $\widehat{M}_i(p)$, we follow Hong and Lee (2005, p.519) to use the $N(0,1)$ cdf truncated on $[-3,3]$ for the weighting function $W(\cdot)$, the Bartlet kernel $K(z) = (1-|z|)I(|z| \leq 1)$ for $K(\cdot)$, and $p = \widehat{p}_0$, where $\widehat{p}_0$ is the data-driven bandwidth via their plug-in method with a preliminary bandwidth $\bar{p} = c(10n)^{1/5}$ for $c = 2, 4, 6$. As Hong and Lee (2005), the resulting $\widehat{M}_i(p)$ is denoted as $\widehat{M}_1(c)$ to highlight the impact from the choice of c.

The tests $D_{n,C}^2$ and $D_{n,I}^2$ are computed as

$$D_{n,C}^2 = \sum_{j=1}^{n} \frac{1}{\widehat{\sigma}_e^2 (j\pi)^2 n_j} \sum_{t=j}^{n}\sum_{s=j}^{n} \widehat{e}_t \widehat{e}_s e^{-\frac{1}{2}(Y_{t-j} - Y_{s-j})^2},$$

$$D_{n,I}^2 = \sum_{j=1}^{n} \frac{1}{\widehat{\sigma}_e^2 (j\pi)^2 n_j n} \sum_{t=1}^{n} \Big\| \sum_{s=j}^{n} \widehat{e}_s I(Y_{s-j} \leq Y_{t-j}) \Big\|^2,$$

where $n_j = n - j + 1$. The critical values of $D^2_{n,C}$ and $D^2_{n,I}$ are obtained by the bootstrap procedure in Escanciano (2006, p.536).

Please insert Table 1 about here.

Table 1 reports the size and power for all examined tests. From Table 1, we have the following findings for the size study:

(i) All MDDM-based tests have a satisfactory size performance in most cases, though they tend to be slightly oversized especially in DGP 2.

(ii) In DGP 1, the tests $D^2_{n,C}$ and $\widehat{M}_i(c)$ $(i = 1, 2, 3)$ are more undersized than our MDDM-based tests. In DGP 2, $\widehat{M}_1(c)$ is slightly undersized, and $D^2_{n,C}$ and $D^2_{n,I}$ tend to be oversized even when the sample size is large, while $\widehat{M}_2(c)$ and $\widehat{M}_3(c)$ suffer a severe oversize problem. The reason for this large size distortion is because $\widehat{M}_2(c)$ and $\widehat{M}_3(c)$ are invalid for the conditional heteroscedastic errors.

Meanwhile, we have the following findings for the power study:

(i) In most examined cases, the power of $\widehat{T}^F_{sn}(M)$ decreases as the value of M increases, while the power of $\widehat{T}^F_{wn}(M)$ is stable over $M = 3, 6, 9$. This is because in our simulation settings, the dependence of residual series usually happens at small lags. Under these settings, $\widehat{T}^F_{sn}(M)$ with a small M (e.g., $M = 3$) can roughly capture all dependence of residual series, and hence they perform best in general. However, the power of $\widehat{T}^F_{sn}(M)$ drops largely as the value of M increases. This is expected, since taking a larger M (e.g., $M = 6$ or 9) for $\widehat{T}^F_{sn}(M)$ can not get more rejection signal but make the critical values larger, in view of the fact that the limiting null distributions of our MDDM tests are stochastically increasing with respect to M. For the tests $\widehat{T}^F_{wn}(M)$, their power performance is robust in terms of M, since the impact of using large M is mitigated by using the weights ω_j.

(ii) For the tests $\widehat{M}_i(c)$ $(i = 1, 2, 3)$, $\widehat{M}_2(c)$ and $\widehat{M}_3(c)$ have a better power performance than $\widehat{M}_1(c)$, which has the size robustness for a wider range of DGPs. The power of all $\widehat{M}_i(c)$ seems mildly sensitive to the choice of c for most DGPs. Similar findings can also be found in Hong and Lee (2005). The tests $D^2_{n,C}$ and $D^2_{n,I}$ are comparable to each other with no dominating ones. The power performance of $\widehat{T}^F_{wn}(M)$ is better than that of $\widehat{T}^F_{sn}(M)$ in most cases, although $\widehat{T}^F_{sn}(3)$ outperforms $\widehat{T}^F_{wn}(3)$ in cases of DGPs 8 and 10. Overall, the pairwise approach used in $\widehat{T}^F_{wn}(M)$ seems preferred due to its stable power performance.

Comparing all the tests, there is no method that outperforms all others in all cases. Our $\widehat{T}^F_{wn}(M)$ appears in the mix of top performers for most examined DGPs.

5.2. *Simulations for $p = 2$.* In this subsection, we consider the simulation studies when the dimension of Y_t is $p = 2$. Our null model is a vector AR(1) (VAR(1)) model:

$$Y_t = A_0 + A_1 Y_{t-1} + \varepsilon_t, \tag{5.2}$$

where $\varepsilon_t = V_t^{1/2}\eta_t$, and $V_t = (v_{t,ij})_{i,j=1,2}$ with

$$\begin{cases} v_{t,11} = \phi_1 + \phi_3 v_{t-1,11} + \phi_5 Y_{t-1,1}^2, \\ v_{t,22} = \phi_2 + \phi_4 v_{t-1,22} + \phi_6 Y_{t-1,2}^2, \\ v_{t,12} = \phi_7 \sqrt{v_{t,11} v_{t,22}}. \end{cases}$$

To examine the size performance of our tests, we generate 1000 replications of sample size n from the following two DGPs based on the model (5.2):

$$\text{DGP 11} : \phi_1 = \phi_2 = 1 \text{ and the others } \phi_3 = \ldots = \phi_7 = 0;$$

$$\text{DGP 12} : \phi_1 = \phi_2 = \phi_5 = \phi_6 = 0.1, \phi_3 = \phi_4 = 0.8 \text{ and } \phi_7 = 0.7,$$

where $A_0 = 0$, $A_1 = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix}$, and η_t is a sequence of i.i.d multivariate normal random variables with mean zero and covariance matrix I_2. Clearly, ε_t is i.i.d. under DGP 11, while it has a multivariate ARCH structure under DGP 12.

To examine the power performance of our tests, we generate 1000 replications of sample size n from the following six DGPs:

$$\text{DGP 13} : Y_t = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix} Y_{t-2} + \varepsilon_t;$$

$$\text{DGP 14} : Y_t = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{pmatrix} \varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP 15} : Y_t = \text{sign}(Y_{t-1}) + 0.43\varepsilon_t, \text{where sign}(x) = I(x > 0) - I(x < 0);$$

$$\text{DGP 16} : Y_{t,j} = \begin{cases} \varrho^{-1} Y_{t-1,j} + \varepsilon_t, & \text{if } 0 \le Y_{t-1,j} < \varrho, \\ (1-\varrho)^{-1}(1 - Y_{t-1,j}) + \varepsilon_t, & \text{if } \varrho \le Y_{t-1,j} \le 1, \end{cases}$$

where $\varrho = 0.49999$, $j = 1, 2$, and each entry of Y_0 follows $U[0, 1]$;

$$\text{DGP 17} : Y_t = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \sin(0.3\pi Y_{t-2}) + \varepsilon_t;$$

$$\text{DGP 18 :} Y_t = \begin{cases} \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} < 0, \\ \begin{pmatrix} -0.6 & 0.4 \\ -0.8 & -0.2 \end{pmatrix} Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} \geq 0, \end{cases}$$

where $\varepsilon_t = \eta_t$. Here, the functionals sign$(\cdot)$ and sin$(\cdot)$ are evaluated elementwisely. With respect to the null VAR(1) model, DGPs 13–14 correspond to the linear models under the alternative, while DGPs 15-18 are the nonlinear models that fall into the alternative.

For each replication, we fit it by model (5.2) and obtain the model residual $\widehat{e}_t$, where $\widehat{e}_t = Y_t - \widehat{A}_{0n} - \widehat{A}_{1n}Y_{t-1}$, and $\widehat{\theta}_n = (vec(\widehat{A}_{0n})', vec(\widehat{A}_{1n})')'$ is the least squares estimate. Based on $\{\widehat{e}_t\}$, we calculate our MDDM-based test statistics $\widehat{T}^F_{sn}(M)$, $\widehat{T}^F_{wn}(M)$, $\widehat{T}^S_{sn}(M)$, and $\widehat{T}^S_{wn}(M)$.

To make a comparison with correlation-based tests, we also compute the portmanteau test statistics $\widehat{Q}_1(M)$ (Box and Pierce (1970)), $\widehat{Q}_2(M)$ (Hosking (1980)), and $\widehat{Q}_3(M)$ (Li and Mcleod (1981)), and the Lagrange multiplier (LM) test statistic $\widehat{LM}(M)$ (Lütkepohl (2005)), where

$$\widehat{Q}_1(M) = n\sum_{i=1}^{M} tr(\widehat{C}_i^T \widehat{C}_0^{-1} \widehat{C}_i \widehat{C}_0^{-1}), \widehat{Q}_2(M) = n\sum_{i=1}^{M} \frac{n}{n-i} tr(\widehat{C}_i^T \widehat{C}_0^{-1} \widehat{C}_i \widehat{C}_0^{-1}),$$
$$\widehat{Q}_3(M) = n\sum_{i=1}^{M} tr(\widehat{C}_i^T \widehat{C}_0^{-1} \widehat{C}_i \widehat{C}_0^{-1}) + \frac{p^2 M(M+1)}{2n}$$

with $\widehat{C}_i = n^{-1}\sum_{t=i+1}^{n} \widehat{e}_t \widehat{e}_{t-i}^T$, and the definition of $\widehat{LM}(M)$ can be found on p.172 of Lütkepohl (2005). At the level α, the critical values of $\widehat{Q}_i(M)$ (i = 1, 2, 3) and $\widehat{LM}(M)$ are $\chi^2_{(M-1)p^2}(\alpha)$ and $\chi^2_{Mp^2}(\alpha)$, respectively, where $\chi^2_s(\alpha)$ is the αth upper percentile of χ^2_s distribution.

Please insert Table 2 about here.

Table 2 reports the size and power of all examined tests. From this table, we find that (i) most MDDM-based tests have a satisfactory size performance for DGP 11, though they tend to be slightly oversized for DGP 12; (ii) all portmanteau tests and LM tests have the similar size performance as our MDDM-based tests for DGP 11, while they suffer a severe oversize problem for DGP 12. The reason for this large size distortion is because the null distributions of all portmanteau tests and the LM test are derived under the assumption that ε_t is i.i.d., which does not hold in the case of DGP 12.

Next, we comment on the findings about the power: (i) In most examined cases, the power of all $\widehat{T}^F_{sn}$, $\widehat{T}^S_{sn}$, $\widehat{Q}_i$, and $\widehat{LM}$ decreases as the value of M increases, while the power of $\widehat{T}^F_{wn}$ and $\widehat{T}^S_{wn}$ is stable over M.

(ii) The MDDM-based tests $\widehat{T}^F_{sn}$ and $\widehat{T}^F_{wn}$ based on Frobenius norm are generally slightly more powerful than their counterparts $\widehat{T}^S_{sn}$ and $\widehat{T}^S_{wn}$ based on spectrum norm and the difference is not substantial in most cases. All portmanteau tests $\widehat{Q}_i$ and the LM test $\widehat{LM}$ have similar power performance.

(iii) For the linear alternative models (i.e., DGPs 13–14), the correlation-based tests $\widehat{Q}_i$ and $\widehat{LM}$ have the best power performance as expected, while the MDDM-based tests $\widehat{T}^F_{wn}$ and $\widehat{T}^S_{wn}$ have inferior power performance when $n = 200$, but their power at $n = 500$ is equal to or close to the best one.

(iv) For the nonlinear alternative models (i.e., DGPs 15–18), all MDDM-based tests in general are much more powerful than the correlation-based tests $\widehat{Q}_i$ and $\widehat{LM}$, especially when the sample size $n = 200$. In particular, the power of $\widehat{Q}_i$ and $\widehat{LM}$ is very low for DGPs 15–16 and 18.

5.3. *Simulations for $p = 5$.* In this subsection, we consider the simulation studies when the dimension of Y_t is $p = 5$. As before, our null model is a VAR(1) model:

$$Y_t = A_0 + A_1 Y_{t-1} + \varepsilon_t, \tag{5.3}$$

where $\varepsilon_t = V_t^{1/2}\eta_t$, and $V_t = (v_{t,ij})_{i,j=1,2,\ldots,5}$ with

$$\begin{cases} v_{t,ii} = \phi_1 + \phi_2 v_{t-1,ii} + \phi_3 Y^2_{t-1,i}, \\ v_{t,ij} = \phi_4\sqrt{v_{t,ii}v_{t,jj}} \quad \text{for} \quad i \neq j. \end{cases}$$

To examine the size performance of our tests, we generate 1000 replications of sample size n from the following two DGPs based on model (5.3):

$$\text{DGP 19}: \phi_1 = 1 \text{ and } \phi_2 = \phi_3 = \phi_4 = 0;$$

$$\text{DGP 20}: \phi_1 = \phi_3 = 0.1, \phi_2 = 0.8 \text{ and } \phi_4 = 0.7,$$

where $A_0 = 0$, $A_1 = 0.3$, and η_t is a sequence of i.i.d multivariate normal random variables with mean zero and covariance matrix I_5. To examine the power performance of our tests, we generate 1000 replications of sample size n from the following six DGPs:

$$\text{DGP 21}: Y_t = 0.3Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t;$$

$$\text{DGP 22}: Y_t = 0.3Y_{t-1} + 0.3\varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP 23}: Y_t = \text{sign}(Y_{t-1}) + 0.43\varepsilon_t;$$

$$\text{DGP 24}: Y_{t,j} = \begin{cases} \varrho^{-1}Y_{t-1,j} + \varepsilon_t, & \text{if } 0 \le Y_{t-1,j} < \varrho, \\ (1-\varrho)^{-1}(1 - Y_{t-1,j}) + \varepsilon_t, & \text{if } \varrho \le Y_{t-1,j} \le 1, \end{cases}$$

where $\varrho = 0.49999$, $j = 1, ..., 5$, and each entry of Y_0 follows $U[0,1]$;

$$\text{DGP 25}: Y_t = 0.3Y_{t-1} + 0.3\sin(0.3\pi Y_{t-2}) + \varepsilon_t;$$

$$\text{DGP 26}: Y_t = \begin{cases} 0.3Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} < 0, \\ -0.3Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} \ge 0, \end{cases}$$

where $\varepsilon_t = \eta_t$. For each replication, we compute all MDDM-based tests and correlation-based tests as done in Subsection 5.2.

Please insert Table 3 about here.

Table 3 reports the size and power of all examined tests. The findings from this table are qualitatively similar to those from Table 2. In terms of size, there is quite a bit size distortion with Frobenius norm-based tests when $n = 200$ for DGP 19, and the size distortion noticeably reduces at sample size $n = 500$. As expected, the correlation-based tests are seriously oversized when ε_t has a multivariate ARCH structure in DGP 20. In terms of power, the MDDM-based tests (especially $\widehat{T}^F_{wn}$) are as competitive as the correlation-based tests to detect linear alternatives, and more importantly, they show clear advantage to detect nonlinear alternatives over the correlation-based tests.

6. Real data example. In this section, we re-study the data set of U.S. monthly interest rates in Tsay (1998). This data set contains the bivariate observations of 3-month treasury bills and 3-year treasury notes from 1959.1 to 1993.2, representing short-term and intermediate series in the term structure of interest rates, respectively. Denote the two interest-rate series by $IR_t = (IR_{1t}, IR_{2t})'$. Let $Y_t = (Y_{1t}, Y_{2t})'$ be the growth series which has 409 observations, where $Y_{it} = \log(IR_{it}) - \log(IR_{i,t-1})$. Tsay (1998) used the following three-regime threshold vector autoregressive (TVAR) model to fit the data set $\{Y_t\}_{t=1}^{409}$, where

$$Y_t = \left[A_0^{(1)} + \sum_{i=1}^{2} A_i^{(1)} Y_{t-i}\right] I(Z_{t-4} \le r_1)$$

$$+\left[A_0^{(2)} + \sum_{i=1}^{6} A_i^{(2)} Y_{t-i}\right] I(r_1 < Z_{t-4} \le r_2) \tag{6.1}$$
$$+\left[A_0^{(3)} + \sum_{i=1}^{7} A_i^{(3)} Y_{t-i}\right] I(Z_{t-4} > r_2) + \varepsilon_t$$

with $r_1 = -0.22817$ and $r_2 = -0.10392$, and the threshold variable Z_t is defined by

$$Z_1 = X_1,\ Z_2 = (X_1 + X_2),\ Z_t = (X_t + X_{t-1} + X_{t-2})/3 \text{ for } t \ge 3,$$

where $X_t = \log(IR_{1t}) - \log(IR_{2t})$ is the 3-month "average spread" in logged interest rates. In model (6.1), the unknown parameters $A_i^{(j)}$ are estimated by the least squares estimates, and the error ε_t has a threshold constant variance structure. According to Assumption 1b in Tsay (1998), a multivariate martingale difference assumption that $E(\varepsilon_t|\mathcal{I}_{t-1}) = 0$ is made but not tested for model (6.1). This motivates us to apply our proposed MDDM-based tests $\widehat{T}_{sn}^F, \widehat{T}_{wn}^F, \widehat{T}_{sn}^S$, and $\widehat{T}_{wn}^S$ to examine the multivariate MDH for ε_t in model (6.1), and the related results are given in Table 4. From this table, we can find the strong evidence from all examined MDDM-based tests that ε_t in model (6.1) is an MDS.

Please insert Table 4 about here.

To make a comparison, we also fit the data set $\{Y_t\}_{t=1}^{409}$ by using a constant model:

$$Y_t = A_0 + \varepsilon_t \tag{6.2}$$

or a VAR(p) model:

$$Y_t = A_0 + \sum_{i=1}^{p} A_i Y_{t-i} + \varepsilon_t, \tag{6.3}$$

where the order p is taken as 2, 6, and 7, which are the autoregressive orders of three regimes in model (6.1). To check whether the constant model and the VAR models can fit $\{Y_t\}_{t=1}^{409}$ adequately, we again apply the MDDM-based tests $\widehat{T}_{sn}^F, \widehat{T}_{wn}^F, \widehat{T}_{sn}^S$, and $\widehat{T}_{wn}^S$ to examine the MDH for ε_t in models (6.2)–(6.3), and the results are given in Table 4. From this table, we can find that the constant model is strongly rejected by all MDDM-based tests, and this indicates that the interest rates market is not efficient. Furthermore, most MDDM-based tests can reject the MDH in all three VAR(p) models at level 5%, and hence the VAR(p) model in (6.3) can not fit $\{Y_t\}_{t=1}^{409}$ adequately. Note that we do not consider the LM test $\widehat{LM}$ and the portmanteau tests $\widehat{Q}_i$ ($i = 1, 2, 3$) for models (6.1)–(6.3), since their errors $\{\varepsilon_t\}$ are not i.i.d.

Overall, our testing results support the use of the TVAR model in (6.1) to fit this bivariate exchange rates data set, which has an unignorable threshold effect as demonstrated from the inadequacy of the constant and linear VAR models.

7. Concluding remarks. In this article, we proposed new MDDM-based tests for the multivariate MDH in stationary time series models, which seems lacking in the literature. Our testing framework is similar to Escanciano (2006), but due to the use of MDDM, we have scale invariance built in for our test statistics. The technical argument is also substantially different from that in the latter paper. Practically speaking, our MDDM-based tests have a wide scope of applicability, as they are robust to high-order dependence of the error term on the past information, as shown in both theoretical and numerical studies. The simulation results and data illustration further lend support to the usefulness of our new tests.

From our extensive simulations, we feel quite comfortable with recommending the use of Frobenious norm-based test coupled with the pairwise approach, that is $\widehat{T}_{wn}^{F}(M)$, whose performance seems quite stable over M. Of course, in practice, how to select M can still be important depending on the nature of linear/nonlinear dependence in the time series. We would recommend the practitioners to try a few different choices of M (or several sets of K_t) to probe possible nonlinear mean dependence. Our proposed bootstrap-based test can be quite conveniently implemented for this purpose.

Finally, we shall point out a few possible future directions. It would be interesting to relax the stationarity assumption and consider model checking for nonstationary time series, and our MDDM-based tests might still be useful. Also, the focus of this paper is on conditional mean independence of the error term given the past information. If the model is constructed for conditional quantiles, then it is natural to extend the proposed test here to test for the conditional quantile independence; see Escanciano and Velasco (2010) for some related work. In addition, model checking for conditional distribution of time series models has been studied in Chen and Hong (2014). It would be interesting to extend the idea of MDDM to that setting.

Acknowledgments. The authors greatly appreciate the very helpful comments and suggestions of two anonymous referees, Associate Editor, and Co-Editor. Wang's research is partially supported by National Social Science Fund of China (no. 20BTJ041), Guangdong

Basic and Applied Basic Research Foundation (no. 2020A1515010821), and Fundamental Research Funds for the Central Universities (no. 12619624). Zhu's research is partially supported by GRF, RGC of Hong Kong (nos. 17306818 and 17305619), NSFC (nos. 11690014 and 11731015), Seed Fund for Basic Research (no. 201811159049), and Fundamental Research Funds for the Central Universities (no. 19JNYH08). Shao's research is partially supported by NSF-DMS (nos. 1807023 and 2014018).

Supplementary materials. The supplementary materials contain some additional simulation results and the Appendix of this paper.

REFERENCES

[1] Bickel, P.J. and Wichura, M.J. (1971) Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Statistics* **42**, 1656-1670.

[2] Bierens, H.J. and Ploberger, W. (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129-1151.

[3] Box, G.E.P. and Pierce, D.A. (1970) Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509-1526.

[4] Chen, B. and Hong, Y. (2014) A unified approach to validating univariate and multivariate conditional distribution models in time series. *Journal of Econometrics* **178**, 22-44.

[5] Chen, S.X. and Gao, J.T. (2007) An adaptive emperical likelihood test for parametric time series regression models. *Journal of Econometrics* **141**, 950-972.

[6] Chen, S.X., Härdle, W. and Li, M. (2003) An emperical likelihood goodness-of-fit for time series. *Journal of the Royal Statistical Society: Series B* **65**, 663-678.

[7] Dahlhaus, T. (2017) Conventional monetary policy transmission during financial crises: an empirical analysis. *Journal of Applied Econometrics* **32**, 401-421.

[8] Delgado, M., Hidalgo, J. and Velasco, C. (2005) Distribution free goodness-of-fit tests for linear models. *Annals of Statistics* **33**, 2568-2609.

[9] Deo, R.S. (2000) Spectral tests of the Martingale Hypothesis under conditional heteroskedasticity. *Journal of Econometrics* **99**, 291-315.

[10] Domínguex, M.A. and Lobato, I.N. (2004) Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72**, 1601-1615.

[11] Durlauf, S. (1991) Spectral-based test for the martingale hypothesis. *Journal of Econometrics* **50**, 1-19.

[12] Escanciano J.C. (2006) Goodness-of-fit tests for linear and non-linear time series models. *Journal of the American Statistical Association* **101**, 531-541.

[13] Escanciano, J.C. (2009) On the lack of power of omnibus specification tests. *Econometric Theory* **25**, 162-194.

[14] Escanciano, J.C. and Velasco, C. (2006) Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* **134**, 151-185.

[15] Escanciano, J.C. and Velasco, C. (2010) Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics* **159**, 209-221.

[16] Gonçalves, S. and Kilian, L. (2004) Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* **123**, 89-120.

[17] Hall, R.E. (1978) Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political Economy* **86**, 971-987.

[18] Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Application*, New York: Academic Press.

[19] Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* **21**, 1926-1947.

[20] Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837-864.

[21] 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.

[22] Hong, Y. and Lee, Y.J. (2005) Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Review of Economic Studies* **72**, 499-541.

[23] Hosking, J.R.M. (1980) The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75**, 602-608.

[24] Khmaladze, E.V. (1981) A martingale approach in the theory of goodness-of-fit tests. *Theory of Probability & Its Applications* **26**, 240-257.

[25] Khmaladze, E.V. (1993) Goodness of fit problem and scanning innovation martingales. *Annals of Statistics* **21**, 798-829.

[26] Khmaladze, E.V. and Koul, H.L. (2004) Martingale transforms goodness-of-fit tests in regression models. *Annals of Statistics* **32**, 995-1034.

[27] Koul, H.L. and Stute, W. (1999) Nonparametric model checks in time series. *Annals of Statistics* **27**, 204-236.

[28] Kuan C.-M. and Lee, W. (2004) A new test of the martingale difference hypothesis. *Studies in Nonlinear Dynamics and Econometrics* **8**, 1-24.

[29] Lee, C.E. and Shao, X. (2018). Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series. *Journal of the American Statistical Association* **113**, 216-229.

[30] Li, W.K. (2003) *Diagnostic Checks in Time Series.* Chapman and Hall/CRC.

[31] Li, W.K. and McLeod, A.I. (1981) Distribution of the residual autocorrelations in multivariate ARMA time series models. *Journal of the Royal Statistical Society: Series B* **43**, 231-239.

[32] Liu, R.Y. (1988) Bootstrap procedure under some non-i.i.d. models. *Annals of Statistics* **16**, 1696-1708.

[33] Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*, Berlin, Springer.

[34] Lo, A.W. (1997). *Market Efficiency: Stock Market Behaviour in Theory and Practice*, vols. I and II. Edward Elgar.

[35] Mammen, E. (1993) Bootstrap and wild bootstrap for high-dimensional linear models. *Annals of Statistics* **21**, 255-285.

[36] Paparoditis, E. (2000) Spectral density based goodness-of-fit tests for time series models. *Scandinavian Journal of Statistics* **27**, 143-176.

[37] Shao, X. (2010) The dependent wild bootstrap. *Journal of the American Statistical Association* **105**, 218-235.

[38] Shao, X. (2011a) A bootstrap-assisted spectral test of white noise under unknown dependence. *Journal of Econometrics* **162**, 213-224.

[39] Shao, X. (2011b) Testing for white noise under unknown dependence and its applications to goodness-of-fit for time series models. *Econometric Theory* **27**, 312-343.

[40] Shao, X. and Zhang, J. (2014) Martingale difference correlation and its use in high dimensional variable screening. *Journal of the American Statistical Association* **109**, 1302-1318.

[41] Stinchcombe, M. and White, H. (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* **14**, 295-325.

[42] Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* **25**, 613-641.

[43] Stute, W., González Manteiga, W. and Presedo Quindimil, M. (1998) Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* **93**, 141-149.

[44] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007) Measuring and testing dependence by correlation of distances. *Annals of Statistics* **35**, 2769-2794.

[45] Teräsvirta, T., Tjøstheim, D. and Granger, C.W.J. (2010) *Modelling Nonlinear Economic Time Series*, Oxford, Oxford University Press.

[46] Teräsvirta, T. and Yang, Y. (2014) Specification, estimation and evaluation of vector smooth transition autoregressive models with applications. *CREATES Research Papers* 2014-08. School of Economics and Management, University of Aarhus.

[47] Tjøstheim, D. (1986) Estimation in nonlinear time series models. *Stochastic Processes and their Applications* **21**, 251-273.

[48] Tripathi, G. and Kitamura, Y. (2003) Testing conditional moment restrictions. *Annals of Statistics* **31**, 2059-2095.

[49] Tsay, R.S. (1998) Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**, 1188-1202.

[50] Whang, Y.J. (2000) Consistent bootstrap tests of parametric regression functions. *Journal of Econometrics* **98**, 27-46.

[51] Whang, Y.J. (2001) Consistent specification testing for conditional moment restrictions. *Economics Letters* **71**, 299-306.

[52] White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.

[53] Wu, C.F.J. (1986) Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* **14**, 1261-1350.

[54] Zhu, K. (2016) Bootstrapping the portmanteau tests in weak auto-regressive moving average models. *Journal of the Royal Statistical Society: Series B* **78**, 463-485.

[55] Zhu, K. and Li, W.K. (2015) A bootstrapped spectral test for adequacy in weak ARMA models. *Journal of Econometrics* **187**, 113-130.

College of Economics
Jinan University
Guangzhou, China
E-mail: twanggc@jnu.edu.cn

Department of Statistics & Actuarial Science
University of Hong Kong
Pok Fu Lam Road, Hong Kong
E-mail: mazhuke@hku.hk

Department of Statistics
University of Illinois at Urbana-Champaign
Champaign, Illinois, USA
E-mail: xshao@illinois.edu

Table 1
The size and power ($\times$100) of all tests for DGPs 1–10 at level 5%

Test \ n	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5	
	200	500	200	500	200	500	200	500	200	500
$\widehat{T}_{sn}^F(3)$	6.0	4.0	5.7	5.5	42.6	70.0	50.6	100	100	100
$\widehat{T}_{sn}^F(6)$	5.7	4.5	5.3	5.3	24.4	46.3	20.3	30.0	99.4	100
$\widehat{T}_{sn}^F(9)$	6.3	6.5	5.4	5.4	20.1	43.5	15.3	23.0	98.4	100
$\widehat{T}_{wn}^F(3)$	4.4	4.5	6.6	6.3	59.6	80.0	96.3	100	100	100
$\widehat{T}_{wn}^F(6)$	4.9	4.9	6.6	6.4	59.5	80.0	96.0	100	100	100
$\widehat{T}_{wn}^F(9)$	4.9	4.9	6.6	6.3	59.4	80.0	96.0	100	100	100
$\widehat{M}_1(2)$	3.0	2.8	3.0	4.3	48.4	78.7	81.3	100	73.2	100
$\widehat{M}_1(4)$	3.8	3.5	2.5	4.1	45.3	74.2	80.1	99.0	70.1	100
$\widehat{M}_1(6)$	3.5	3.6	3.0	3.2	35.2	75.1	79.4	99.3	74.0	100
$\widehat{M}_2(2)$	3.1	2.9	20.1	23.4	75.2	100	88.7	100	90.5	100
$\widehat{M}_2(4)$	3.7	3.8	18.0	22.4	74.3	99.0	86.7	100	89.9	100
$\widehat{M}_2(6)$	4.1	4.2	15.0	18.6	72.2	94.0	82.5	100	90.1	100
$\widehat{M}_3(2)$	3.0	2.7	20.8	22.0	72.0	100	92.3	100	83.4	100
$\widehat{M}_3(4)$	4.1	4.1	20.1	26.0	70.1	99.2	92.4	100	82.3	100
$\widehat{M}_3(6)$	3.8	4.0	19.8	25.0	70.4	86.1	81.6	100	80.7	100
$D_{n,C}^2$	4.1	3.0	6.3	10.0	68.0	87.0	96.5	100	100	100
$D_{n,I}^2$	4.7	4.5	5.6	7.0	52.5	74.0	92.9	100	100	100

Test \ n	DGP 6		DGP 7		DGP 8		DGP 9		DGP 10	
	200	500	200	500	200	500	200	500	200	500
$\widehat{T}_{sn}^F(3)$	30.0	90.0	100	100	95.6	100	56.0	100	41.0	43.0
$\widehat{T}_{sn}^F(6)$	19.3	30.0	100	100	76.0	100	31.0	54.0	100	100
$\widehat{T}_{sn}^F(9)$	17.0	18.0	99.0	100	64.9	98.0	22.0	28.0	100	100
$\widehat{T}_{wn}^F(3)$	92.0	100	100	100	80.6	100	89.0	100	26.8	27.0
$\widehat{T}_{wn}^F(6)$	92.0	100	100	100	80.6	100	89.0	100	95.3	100
$\widehat{T}_{wn}^F(9)$	91.4	100	100	100	81.1	100	90.0	100	95.2	100
$\widehat{M}_1(2)$	54.3	90.7	80.3	100	85.7	100	87.6	100	80.4	100
$\widehat{M}_1(4)$	57.2	88.3	69.2	90.0	75.3	96.0	84.4	100	92.1	100
$\widehat{M}_1(6)$	49.5	87.5	68.5	89.8	69.8	95.4	81.3	100	89.5	100
$\widehat{M}_2(2)$	65.2	93.9	82.3	100	85.6	100	75.2	98.2	84.3	100
$\widehat{M}_2(4)$	65.3	90.5	81.0	100	75.9	95.6	86.4	100	100	100
$\widehat{M}_2(6)$	64.9	91.9	75.9	98.4	68.8	93.2	82.8	100	100	100
$\widehat{M}_3(2)$	69.5	92.6	83.3	100	91.3	100	91.0	100	83.0	100
$\widehat{M}_3(4)$	60.8	90.0	82.5	100	87.7	100	85.7	100	95.4	100
$\widehat{M}_3(6)$	56.5	87.3	75.3	98.3	80.4	97.8	83.5	100	97.4	100
$D_{n,C}^2$	93.7	100	100	100	23.9	68.0	88.0	100	70.2	100
$D_{n,I}^2$	85.6	100	100	100	80.4	100	84.0	99.0	99.3	100

Table 2
The size and power (×100) of all tests for DGPs 11–18 at level 5%.

Test \ n	DGP 11		DGP 12		DGP 13		DGP 14	
	200	500	200	500	200	500	200	500
$\widehat{T}^F_{sn}(3)$	5.3	5.2	6.1	6.0	100	100	95.0	100
$\widehat{T}^F_{sn}(6)$	6.9	5.4	8.1	6.0	98.8	100	63.0	99.6
$\widehat{T}^F_{sn}(9)$	7.1	4.6	7.2	4.7	90.5	100	50.1	94.9
$\widehat{T}^S_{sn}(3)$	5.2	4.9	7.1	7.0	100	100	93.8	100
$\widehat{T}^S_{sn}(6)$	6.2	5.8	8.0	6.0	98.4	100	60.3	99.4
$\widehat{T}^S_{sn}(9)$	6.4	4.6	7.1	4.7	87.5	100	49.1	94.7
$\widehat{T}^F_{wn}(3)$	6.0	5.9	7.6	7.5	99.5	100	77.4	100
$\widehat{T}^F_{wn}(6)$	6.4	5.9	7.9	7.7	99.5	100	77.8	100
$\widehat{T}^F_{wn}(9)$	6.4	5.7	7.3	7.2	99.5	100	78.6	100
$\widehat{T}^S_{wn}(3)$	5.5	5.3	7.5	7.4	99.5	100	75.7	100
$\widehat{T}^S_{wn}(6)$	6.2	5.8	7.2	7.1	99.5	100	76.4	100
$\widehat{T}^S_{wn}(9)$	6.1	5.6	7.4	7.3	99.5	100	76.9	100
$\widehat{Q}_1(3)$	5.9	5.8	24.5	35.4	100	100	100	100
$\widehat{Q}_1(6)$	6.1	4.9	33.2	46.2	100	100	100	100
$\widehat{Q}_1(9)$	5.4	5.7	38.5	54.0	100	100	99.7	100
$\widehat{Q}_2(3)$	5.9	6.0	25.1	35.5	100	100	100	100
$\widehat{Q}_2(6)$	7.0	5.3	35.5	46.9	100	100	100	100
$\widehat{Q}_2(9)$	7.1	6.3	41.9	54.9	100	100	99.9	100
$\widehat{Q}_3(3)$	5.9	5.8	25.0	35.4	100	100	100	100
$\widehat{Q}_3(6)$	6.8	5.3	35.0	46.5	100	100	100	100
$\widehat{Q}_3(9)$	6.9	6.2	40.9	54.7	100	100	99.9	100
$\widehat{LM}(3)$	5.3	4.8	29.2	39.9	100	100	100	100
$\widehat{LM}(6)$	4.7	5.0	33.3	49.7	100	100	100	100
$\widehat{LM}(9)$	3.6	4.9	33.5	54.6	100	100	100	100
	DGP 15		DGP 16		DGP 17		DGP 18	
Test \ n	200	500	200	500	200	500	200	500
$\widehat{T}^F_{sn}(3)$	25.5	73.0	14.5	34.7	96.1	100	67.7	99.8
$\widehat{T}^F_{sn}(6)$	10.0	33.0	12.1	19.5	51.7	98.3	27.5	72.7
$\widehat{T}^F_{sn}(9)$	11.0	27.0	7.5	12.4	27.5	84.7	17.1	46.0
$\widehat{T}^S_{sn}(3)$	21.0	64.0	13.5	30.5	95.4	100	61.0	99.6
$\widehat{T}^S_{sn}(6)$	10.9	33.6	10.9	16.4	50.0	98.5	24.9	67.9
$\widehat{T}^S_{sn}(9)$	11.4	26.8	9.8	13.2	27.1	84.2	16.3	42.9
$\widehat{T}^F_{wn}(3)$	99.0	100	99.5	100	98.0	100	100	100
$\widehat{T}^F_{wn}(6)$	99.5	100	99.5	100	97.4	100	100	100
$\widehat{T}^F_{wn}(9)$	99.0	100	99.5	100	97.4	100	100	100
$\widehat{T}^S_{wn}(3)$	97.5	100	98.0	100	97.9	100	100	100
$\widehat{T}^S_{wn}(6)$	96.0	100	97.5	100	97.4	100	100	100
$\widehat{T}^S_{wn}(9)$	96.5	100	97.5	100	97.3	100	100	100
$\widehat{Q}_1(3)$	18.1	42.5	6.1	7.9	70.4	98.7	14.0	27.0
$\widehat{Q}_1(6)$	14.5	21.5	5.7	12.1	59.6	96.9	10.1	20.2
$\widehat{Q}_1(9)$	13.2	27.3	6.6	7.1	50.7	93.1	8.1	17.7
$\widehat{Q}_2(3)$	19.5	42.0	6.6	8.5	70.8	98.7	14.5	27.4
$\widehat{Q}_2(6)$	16.5	30.2	6.1	13.3	61.6	97.0	11.4	21.0
$\widehat{Q}_2(9)$	16.0	28.6	7.6	9.5	53.7	93.9	10.6	18.5
$\widehat{Q}_3(3)$	20.7	42.8	6.5	7.9	70.8	98.7	14.4	27.4
$\widehat{Q}_3(6)$	15.8	30.9	6.0	13.4	61.3	97.0	10.8	21.0
$\widehat{Q}_3(9)$	14.6	29.3	7.1	8.4	53.2	93.9	9.3	18.5
$\widehat{LM}(3)$	15.6	37.8	14.0	15.6	67.4	98.7	7.8	19.5
$\widehat{LM}(6)$	11.3	29.6	8.5	15.3	47.5	96.1	6.4	15.8
$\widehat{LM}(9)$	8.4	21.1	6.0	11.6	35.6	89.9	5.3	13.1

Table 3
The size and power (×100) of all tests for DGPs 19–26 at level 5%.

Test \ n	DGP 19 200	DGP 19 500	DGP 20 200	DGP 20 500	DGP 21 200	DGP 21 500	DGP 22 200	DGP 22 500
$\widehat{T}^F_{sn}(3)$	8.9	8.1	6.3	5.6	88.5	100	68.4	97.0
$\widehat{T}^F_{sn}(6)$	10.6	7.3	7.1	5.9	70.8	99.7	53.2	83.8
$\widehat{T}^F_{sn}(9)$	12.2	8.8	7.4	5.8	65.6	98.9	50.5	74.0
$\widehat{T}^S_{sn}(3)$	7.4	7.2	6.3	5.6	70.9	99.6	50.5	85.9
$\widehat{T}^S_{sn}(6)$	6.9	6.0	7.0	5.9	48.4	92.7	30.1	52.1
$\widehat{T}^S_{sn}(9)$	6.9	6.7	7.3	5.8	41.4	83.2	25.6	41.0
$\widehat{T}^F_{wn}(3)$	11.8	7.1	6.0	5.2	92.4	100	73.3	97.6
$\widehat{T}^F_{wn}(6)$	12.8	7.0	6.1	5.7	92.7	100	72.7	97.4
$\widehat{T}^F_{wn}(9)$	12.5	7.4	6.3	5.4	92.4	100	73.8	97.2
$\widehat{T}^S_{wn}(3)$	8.8	6.7	6.0	5.2	77.2	99.4	55.7	89.4
$\widehat{T}^S_{wn}(6)$	9.0	6.5	6.0	5.7	78.2	99.5	56.7	88.9
$\widehat{T}^S_{wn}(9)$	8.9	6.7	6.4	5.6	78.4	99.5	56.4	88.6
$\widehat{Q}_1(3)$	6.5	5.6	16.4	17.9	84.4	100	68.1	99.5
$\widehat{Q}_1(6)$	5.8	5.8	23.4	22.4	69.4	100	52.9	94.0
$\widehat{Q}_1(9)$	6.8	6.5	29.5	28.8	65.3	99.7	50.1	88.4
$\widehat{Q}_2(3)$	7.2	6.1	17.9	18.2	85.1	100	70.6	99.5
$\widehat{Q}_2(6)$	8.6	6.9	28.4	24.3	74.2	100	59.5	94.5
$\widehat{Q}_2(9)$	11.4	8.8	39.4	32.0	74.5	99.8	59.7	90.5
$\widehat{Q}_3(3)$	7.1	6.0	17.7	18.2	85.2	100	70.6	99.5
$\widehat{Q}_3(6)$	8.2	6.7	27.3	24.1	74.3	100	59.1	94.5
$\widehat{Q}_3(9)$	10.6	8.4	38.2	31.4	74.1	99.8	58.9	90.5
$\widehat{LM}(3)$	3.8	5.2	13.1	18.3	65.4	100	58.2	98.8
$\widehat{LM}(6)$	5.3	4.3	11.8	15.1	41.0	99.4	36.7	91.2
$\widehat{LM}(9)$	4.7	4.9	10.3	12.8	30.1	97.3	24.7	80.2

Test \ n	DGP 23 200	DGP 23 500	DGP 24 200	DGP 24 500	DGP 25 200	DGP 25 500	DGP 26 200	DGP 26 500
$\widehat{T}^F_{sn}(3)$	30.5	65.0	14.5	18.7	71.9	99.7	9.7	9.8
$\widehat{T}^F_{sn}(6)$	22.9	25.5	14.5	15.6	52.9	94.5	10.5	11.0
$\widehat{T}^F_{sn}(9)$	20.4	23.6	17.5	20.0	50.3	89.4	13.4	14.0
$\widehat{T}^S_{sn}(3)$	21.9	45.0	10.6	12.8	54.8	95.3	7.5	8.0
$\widehat{T}^S_{sn}(6)$	12.3	16.4	10.8	12.6	36.2	75.7	6.5	7.0
$\widehat{T}^S_{sn}(9)$	14.3	15.3	9.2	12.3	31.4	61.6	7.6	8.0
$\widehat{T}^F_{wn}(3)$	76.8	100	82.3	100	79.1	99.5	56.6	96.0
$\widehat{T}^F_{wn}(6)$	73.6	100	78.7	100	78.9	99.5	55.7	96.0
$\widehat{T}^F_{wn}(9)$	76.0	99.0	78.5	100	79.5	99.5	57.5	95.0
$\widehat{T}^S_{wn}(3)$	56.0	89.0	60.6	95.0	62.7	96.7	36.4	78.0
$\widehat{T}^S_{wn}(6)$	53.5	86.0	60.8	94.4	61.5	97.4	36.2	78.0
$\widehat{T}^S_{wn}(9)$	54.5	87.0	59.7	94.1	62.4	97.2	36.9	76.0
$\widehat{Q}_1(3)$	22.7	53.0	5.2	6.9	61.4	99.8	4.9	5.0
$\widehat{Q}_1(6)$	18.8	34.0	6.1	7.1	48.8	97.3	6.1	6.2
$\widehat{Q}_1(9)$	19.7	32.0	6.0	8.1	46.3	93.3	7.0	7.1
$\widehat{Q}_2(3)$	24.7	54.7	6.1	7.5	64.0	99.8	6.0	6.2
$\widehat{Q}_2(6)$	20.0	35.2	10.2	12.5	54.4	97.6	8.2	8.3
$\widehat{Q}_2(9)$	33.4	36.2	12.5	13.6	55.2	95.0	12.4	12.5
$\widehat{Q}_3(3)$	24.7	54.6	6.0	7.5	64.0	99.8	5.9	6.0
$\widehat{Q}_3(6)$	19.5	35.1	10.3	12.4	54.1	97.6	7.8	8.0
$\widehat{Q}_3(9)$	30.9	36.3	12.5	13.3	54.0	95.1	11.2	12.0
$\widehat{LM}(3)$	17.9	46.0	14.2	19.3	38.0	99.1	3.4	3.8
$\widehat{LM}(6)$	10.6	28.0	12.4	16.1	25.4	89.9	4.1	4.2
$\widehat{LM}(9)$	8.2	23.0	8.3	9.7	19.0	78.6	5.1	5.2

TABLE 4
The p-values of four MDDM-based tests at lag $M = 1, ..., 10$ for five different models

		Tests						Tests			
Model	M	$\widehat{T}_{sn}^F$	$\widehat{T}_{sn}^S$	$\widehat{T}_{wn}^F$	$\widehat{T}_{wn}^S$	Model	M	$\widehat{T}_{sn}^F$	$\widehat{T}_{sn}^S$	$\widehat{T}_{wn}^F$	$\widehat{T}_{wn}^S$
TVAR	1	**0.641**	**0.640**	**0.641**	**0.640**	Constant	1	0.000	0.000	0.000	0.000
	2	**0.838**	**0.834**	**0.718**	**0.709**		2	0.000	0.000	0.000	0.000
	3	**0.766**	**0.758**	**0.717**	**0.712**		3	0.000	0.000	0.000	0.000
	4	**0.833**	**0.830**	**0.729**	**0.725**		4	0.000	0.000	0.000	0.000
	5	**0.869**	**0.865**	**0.733**	**0.733**		5	0.000	0.000	0.000	0.000
	6	**0.914**	**0.913**	**0.762**	**0.755**		6	0.000	0.000	0.000	0.000
	7	**0.908**	**0.904**	**0.762**	**0.754**		7	0.000	0.000	0.000	0.000
	8	**0.894**	**0.891**	**0.762**	**0.754**		8	0.000	0.000	0.000	0.000
	9	**0.888**	**0.887**	**0.762**	**0.754**		9	0.000	0.000	0.000	0.000
	10	**0.907**	**0.904**	**0.762**	**0.759**		10	0.000	0.000	0.000	0.000
VAR(2)	1	0.047	0.047	0.047	0.047	VAR(6)	1	**0.051**	**0.052**	**0.051**	**0.052**
	2	0.023	0.023	0.029	0.029		2	0.020	0.020	0.031	0.031
	3	0.025	0.025	0.024	0.024		3	0.005	0.005	0.028	0.028
	4	**0.070**	**0.071**	0.025	0.025		4	0.003	0.003	0.027	0.027
	5	**0.126**	**0.126**	0.024	0.024		5	0.009	0.009	0.023	0.023
	6	0.020	0.020	0.016	0.016		6	0.007	0.007	0.023	0.023
	7	0.019	0.020	0.015	0.015		7	0.016	0.017	0.024	0.024
	8	0.023	0.025	0.017	0.016		8	0.022	0.024	0.024	0.024
	9	0.043	0.044	0.015	0.015		9	0.033	0.034	0.024	0.024
	10	**0.062**	**0.062**	0.015	0.015		10	**0.075**	**0.075**	0.024	0.024
VAR(7)	1	**0.066**	**0.066**	**0.066**	**0.066**						
	2	0.014	0.014	0.026	0.026						
	3	0.002	0.002	0.023	0.023						
	4	0.004	0.004	0.022	0.022						
	5	0.008	0.008	0.019	0.019						
	6	0.012	0.012	0.020	0.020						
	7	0.008	0.008	0.020	0.020						
	8	0.031	0.031	0.019	0.019						
	9	0.039	0.040	0.019	0.018						
	10	0.023	0.023	0.019	0.019						

† The p-value larger than 5% is in boldface.

SUPPLEMENTARY MATERIALS FOR "TESTING FOR THE MARTINGALE DIFFERENCE HYPOTHESIS IN MULTIVARIATE TIME SERIES MODELS"

By Guochang Wang*, Ke Zhu† and Xiaofeng Shao‡

Jinan University, University of Hong Kong†*

and University of Illinois at Urbana-Champaign‡

This supplement provides some additional simulation results and the Appendix of the paper.

Additional Simulations

In this section, we first give some additional simulations to check how our proposed MDDM-based tests perform under persistent GARCH or Stochastic volatility (SV) errors, based on the following five DGPs:

DGP 27: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t$

DGP 28: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t\sigma_t$ and $\sigma_t^2 = 0.001 + 0.01\varepsilon_{t-1}^2 + 0.97\sigma_{t-1}^2$;

DGP 29: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t\sigma_t$ and $\sigma_t^2 = 0.001 + 0.09\varepsilon_{t-1}^2 + 0.89\sigma_{t-1}^2$;

DGP 30: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t\sigma_t$ and $\sigma_t^2 = 0.001 + 0.09\varepsilon_{t-1}^2 + 0.90\sigma_{t-1}^2$;

DGP 31: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t \exp(\sigma_t)$ and $\sigma_t = 0.936\sigma_{t-1} + 0.32u_t$,

where η_t and u_t are independent sequences of i.i.d. $N(0,1)$. Note that DGPs 27–31 correspond to the case that the conditional mean of Y_t is a constant. Particularly, DGPs 28–31 are just the designs 2 and 3 in Escanciano and Velasco (2006) to mimic the persistent GARCH or Stochastic volatility (SV) errors found in many financial applications (see, e.g., O'Hara et al. (2014)), and our moment conditions in Assumption 3.1 are violated under these DGPs.

Panel A of Table 0.1 reports the size results of our MDDM-based tests $\widehat{T}_{sn}^F(M)$ and $\widehat{T}_{wn}^F(M)$ (for $M = 3, 6, 9$) and the tests $D_{n,C}^2$ and $D_{n,I}^2$ in Escanciano (2006) under DGPs 27–31. From this table, we can find that all considered tests have a precise size performance. This finding suggests that the moment conditions in Assumption 3.1 are perhaps only a theoretical matter, and our proposed MDDM-based tests can still be used under persistent GARCH or SV errors with an unbounded higher moment.

TABLE 0.1
The size (in percentage) of all tests for DGPs 27–31 at level 5%

Test \ n	DGP 27		DGP 28		DGP 29		DGP 30		DGP 31	
	200	500	200	500	200	500	200	500	200	500
					Panel A: The results for Y_t					
$\widehat{T}_{sn}^F(3)$	2.5	5.5	6.5	6.3	6.5	6.5	5.5	5.4	3.5	5.0
$\widehat{T}_{sn}^F(6)$	5.5	4.8	6.5	6.3	6.5	6.5	6.5	6.5	2.5	5.0
$\widehat{T}_{sn}^F(9)$	6.0	5.8	6.5	6.4	6.0	6.0	5.5	5.3	4.0	4.5
$\widehat{T}_{wn}^F(3)$	5.5	5.3	6.5	6.5	7.5	6.5	6.5	6.4	6.0	4.0
$\widehat{T}_{wn}^F(6)$	4.0	5.5	6.0	5.9	7.0	7.0	6.0	5.8	5.5	4.7
$\widehat{T}_{wn}^F(9)$	4.0	5.4	6.5	6.5	6.5	6.5	6.5	6.3	5.5	4.8
$D_{n,C}^2$	3.5	6.0	6.0	6.0	6.0	6.0	6.0	6.0	3.0	4.0
$D_{n,I}^2$	4.5	8.0	6.0	6.2	6.0	5.5	6.5	5.5	5.0	4.0
					Panel B: The results for $100Y_t$					
$\widehat{T}_{sn}^F(3)$	2.5	5.5	6.5	6.3	6.5	6.5	5.5	5.4	3.5	5.0
$\widehat{T}_{sn}^F(6)$	5.5	4.8	6.5	6.3	6.5	6.5	6.5	6.5	2.5	5.0
$\widehat{T}_{sn}^F(9)$	6.0	5.8	6.5	6.4	6.0	6.0	5.5	5.3	4.0	4.5
$\widehat{T}_{wn}^F(3)$	5.5	5.3	6.5	6.5	7.5	6.5	6.5	6.4	6.0	4.0
$\widehat{T}_{wn}^F(6)$	4.0	5.5	6.0	5.9	7.0	7.0	6.0	5.8	5.5	4.7
$\widehat{T}_{wn}^F(9)$	4.0	5.4	6.5	6.5	6.5	6.5	6.5	6.3	5.5	4.8
$D_{n,C}^2$	3.0	6.5	7.0	6.4	7.0	5.0	7.5	5.5	4.5	4.5
$D_{n,I}^2$	4.5	8.0	6.0	6.2	6.0	5.5	6.5	5.5	5.0	4.0
					Panel C: The results for $Y_t/100$					
$\widehat{T}_{sn}^F(3)$	2.5	5.5	6.5	6.3	6.5	6.5	5.5	5.4	3.5	5.0
$\widehat{T}_{sn}^F(6)$	5.5	4.8	6.5	6.3	6.5	6.5	6.5	6.5	2.5	5.0
$\widehat{T}_{sn}^F(9)$	6.0	5.8	6.5	6.4	6.0	6.0	5.5	5.3	4.0	4.5
$\widehat{T}_{wn}^F(3)$	5.5	5.3	6.5	6.5	7.5	6.5	6.5	6.4	6.0	4.0
$\widehat{T}_{wn}^F(6)$	4.0	5.5	6.0	5.9	7.0	7.0	6.0	5.8	5.5	4.7
$\widehat{T}_{wn}^F(9)$	4.0	5.4	6.5	6.5	6.5	6.5	6.5	6.3	5.5	4.8
$D_{n,C}^2$	14.5	15.5	5.0	5.0	5.0	5.0	5.0	4.5	7.0	8.0
$D_{n,I}^2$	4.5	8.0	6.0	6.2	6.0	5.5	6.5	5.5	5.0	4.0

Next, we give some additional simulations to check the scale invariance of our MDDM-based tests and the tests $D_{n,C}^2$ and $D_{n,I}^2$ in Escanciano (2006). For the size study, we consider the DGPs 27–31 as before. For the power study, we also follow Escanciano and Velasco (2006) to consider the following DGPs:

DGP 32: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t + X_t - X_{t-1}$ and $X_t = 0.85X_{t-1} + u_t$;

DGP 33: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_t + 0.25\eta_{t-1}Y_{t-1} + 0.15\eta_{t-1}Y_{t-2}$;

DGP 34: $Y_t = \varepsilon_t$ with $\varepsilon_t = \eta_{t-1}\eta_{t-2}(\eta_{t-2} + \eta_t + 1)$;

DGP 35: $Y_t = \varepsilon_t$ with $\varepsilon_t = -0.5\varepsilon_{t-1} + \eta_t$ if $\varepsilon_{t-1} \geq 1,$ and $\varepsilon_t = 0.4\varepsilon_{t-1} + \eta_t,$ if $\varepsilon_{t-1} < 1$;

DGP 36: $Y_t = \varepsilon_t$ with $\varepsilon_t = 0.6\varepsilon_{t-1}\exp(-0.5\varepsilon_{t-1}^2) + \eta_t$,

TABLE 0.2
The power (in percentage) of all tests for DGPs 32–36 at level 5%

Test \ n	DGP 32		DGP 33		DGP 34		DGP 35		DGP 36	
	200	500	200	500	200	500	200	500	200	500
Panel A: The results for Y_t										
$\widehat{T}_{sn}^F(3)$	8.5	14.0	29.0	86.5	12.0	15.5	53.0	98.5	70.5	99.5
$\widehat{T}_{sn}^F(6)$	8.0	14.5	18.0	50.4	12.5	13.0	32.5	77.5	55.5	96.5
$\widehat{T}_{sn}^F(9)$	6.5	13.5	14.5	34.5	13.0	14.0	26.0	62.5	42.5	93.6
$\widehat{T}_{wn}^F(3)$	8.5	12.0	73.0	100	19.0	34.5	90.5	100	90.5	100
$\widehat{T}_{wn}^F(6)$	8.5	13.0	71.5	100	18.0	35.5	89.0	100	90.5	100
$\widehat{T}_{wn}^F(9)$	8.5	13.5	71.5	100	18.0	36.0	90.0	100	90.5	100
$D_{n,C}^2$	6.5	9.0	91.0	100	29.0	65.0	88.0	100	89.5	100
$D_{n,I}^2$	7.5	11.0	51.0	99.0	32.0	51.5	81.5	100	88.5	100
Panel B: The results for $100Y_t$										
$\widehat{T}_{sn}^F(3)$	8.5	14.0	29.0	86.5	12.0	15.5	53.0	98.5	70.5	99.5
$\widehat{T}_{sn}^F(6)$	8.0	14.5	18.0	50.4	12.5	13.0	32.5	77.5	55.5	96.5
$\widehat{T}_{sn}^F(9)$	6.5	13.5	14.5	34.5	13.0	14.0	26.0	62.5	42.5	93.6
$\widehat{T}_{wn}^F(3)$	8.5	12.0	73.0	100	19.0	34.5	90.5	100	90.5	100
$\widehat{T}_{wn}^F(6)$	8.5	13.0	71.5	100	18.0	35.5	89.0	100	90.5	100
$\widehat{T}_{wn}^F(9)$	8.5	13.5	71.5	100	18.0	36.0	90.0	100	90.5	100
$D_{n,C}^2$	4.5	4.0	20.5	44.5	12.5	31.0	31.0	82.5	24.5	57.5
$D_{n,I}^2$	7.5	11.0	51.0	99.0	32.0	51.5	81.5	100	88.5	100
Panel C: The results for $Y_t/100$										
$\widehat{T}_{sn}^F(3)$	8.5	14.0	29.0	86.5	12.0	15.5	53.0	98.5	70.5	99.5
$\widehat{T}_{sn}^F(6)$	8.0	14.5	18.0	50.4	12.5	13.0	32.5	77.5	55.5	96.5
$\widehat{T}_{sn}^F(9)$	6.5	13.5	14.5	34.5	13.0	14.0	26.0	62.5	42.5	93.6
$\widehat{T}_{wn}^F(3)$	8.5	12.0	73.0	100	19.0	34.5	90.5	100	90.5	100
$\widehat{T}_{wn}^F(6)$	8.5	13.0	71.5	100	18.0	35.5	89.0	100	90.5	100
$\widehat{T}_{wn}^F(9)$	8.5	13.5	71.5	100	18.0	36.0	90.0	100	90.5	100
$D_{n,C}^2$	0.5	0.5	10.5	9.5	6.0	4.0	8.5	14.5	13.0	19.0
$D_{n,I}^2$	7.5	11.0	51.0	99.0	32.0	51.5	81.5	100	88.5	100

where η_t and u_t are independent sequences of i.i.d. $N(0,1)$. Tables 0.1 and 0.2 report the size and power of all considered tests, based on the original data Y_t and the re-scaled data $100Y_t$ or $Y_t/100$. From these two tables, we can clearly see that as expected, the results of our MDDM-based tests and the test $D_{n,I}^2$ are scale-invariant, while the results of the test $D_{n,C}^2$ are not.

APPENDIX: PROOFS

PROOF OF EQUATION (2.1). Recall $V = (V_1, ..., V_p)^T \in \mathcal{R}^p$, $U \in \mathcal{R}^q$, and (V', U') is an i.i.d. copy of (V, U). For any $j, k = 1, ..., p$, by the independence of (V, U) and (V', U'), we have

$$\begin{aligned} \mathcal{G}_j(s)\mathcal{G}_k(s)^\star &= E\left[(V_j - E(V_j))\, e^{i\langle s,U\rangle}\right] E\left[\left(V_k' - E(V_k')\right) e^{-i\langle s,U'\rangle}\right] \\ &= E\left[(V_j - E(V_j))\left(V_k' - E(V_k')\right) e^{i\langle s,U-U'\rangle}\right] \end{aligned}$$

$$
\begin{aligned}
&= E\left[(V_j - E(V_j))\left(V_k' - E(V_k')\right)\left\{\cos(\langle s, U-U'\rangle) + i\sin(\langle s, U-U'\rangle)\right\}\right] \\
&= -E\left[(V_j - E(V_j))\left(V_k' - E(V_k')\right)\{1-\cos(\langle s, U-U'\rangle)\}\right] \\
&+ iE\left[(V_j - E(V_j))\left(V_k' - E(V_k')\right)\sin(\langle s, U-U'\rangle)\right] \\
&=: A_{jk}(s) + B_{jk}(s).
\end{aligned}
$$

For any $j, k = 1, ..., p$, since $B_{jk}(s)$ is an odd function, we have

$$
\frac{1}{c_q}\int_{\mathcal{R}^q} \frac{B_{jk}(s)}{\|s\|^{1+q}} ds = 0,
$$

and hence the result (2.1) follows directly from the fact that

$$
\begin{aligned}
&\frac{1}{c_q}\int_{\mathcal{R}^q} \frac{\mathcal{G}_j(s)\mathcal{G}_k(s)^\star}{\|s\|^{1+q}} ds \\
&= \frac{1}{c_q}\int_{\mathcal{R}^q} \frac{A_{jk}(s)}{\|s\|^{1+q}} ds \\
&= -\frac{1}{c_q}E\left[(V_j - E(V_j))\left(V_k' - E(V_k')\right)\int_{\mathcal{R}^q} \frac{1-\cos(\langle s, U-U'\rangle)}{\|s\|^{1+q}} ds\right] \\
&= -E\left[(V_j - E(V_j))\left(V_k' - E(V_k')\right)\|U-U'\|\right],
\end{aligned}
$$

where the last equality holds since by Lemma 1 in Székely and Rizzo (2005),

$$
\int_{\mathcal{R}^q} \frac{1-\cos(\langle s, x\rangle)}{\|s\|^{1+q}} ds = c_q\|x\|
$$

for all $x \in \mathcal{R}^q$. □

Let $\widetilde{e}_t = Y_t - f(\mathcal{I}_{t-1}, \widehat{\theta}_n)$. Define $\widetilde{\mathcal{G}}_n(s)$, $\mathcal{G}_n(s)$, and $\mathcal{G}_{n,\varepsilon}(s)$ in the same way as $\widehat{\mathcal{G}}_n(s)$ in (3.3) with $\widehat{e}_t$ replaced by $\widetilde{e}_t$, e_t, and ε_t, respectively. To facilitate the proof of Theorem 3.1, we need five technical lemmas. Lemma A.1 gives a useful inequality. Lemma A.2 shows that replacing $\widehat{\mathcal{G}}_n(s)$ by $\widetilde{\mathcal{G}}_n(s)$ has no effect on the asymptotic theory of our test statistic $\widehat{T}_n^F$. Lemma A.3 provides a useful expansion on $\widetilde{\mathcal{G}}_n(s)$. Lemma A.4 proves the weak convergence of $\mathcal{G}_{n,\varepsilon}(s)$. Lemma A.5 offers the joint convergence of $\mathcal{G}_{n,\varepsilon}(s)$ and $\widehat{\theta}_n$.

LEMMA A.1. *For any $s \in \mathcal{R}^k$,*

$$
|\cos\langle s, K_t\rangle - 1| \le C\left(1 \wedge (\|s\|\|K_t\|)^2\right). \tag{A.1}
$$

PROOF. The conclusion holds since $|\cos\langle s, K_t\rangle - 1| = 2\big(\sin\frac{\langle s, K_t\rangle}{2}\big)^2$. □

LEMMA A.2. *Suppose that Assumptions 3.1(i), (iii) and 3.5 hold. Then,*

(i) $N \cdot E|\widehat{\mathcal{G}}_n(s) - \widetilde{\mathcal{G}}_n(s)|_p^2 = o\left(1 \wedge \|s\|^2\right)$,

where $o(1)$ *holds uniformly in* $\{s : s \in \mathcal{R}^k\}$*; consequently,*

(ii) $N \cdot E|\widehat{\mathcal{G}}_n(s) - \widetilde{\mathcal{G}}_n(s)|_p^2 = o(1)$

holds uniformly over any compact set $\Omega \subset \mathcal{R}^k$*.*

PROOF. It suffices to prove (i). By Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
&|\widehat{\mathcal{G}}_n(s) - \widetilde{\mathcal{G}}_n(s)|_p^2 \\
&\leq 2\left|\frac{1}{N}\sum_{t=n_0}^{n}[f(\mathcal{I}_{t-1},\widehat{\theta}_n) - f(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)](e^{i\langle s,K_t\rangle}-1)\right|_p^2 \\
&\quad + 2\left|\frac{1}{N}\sum_{t=n_0}^{n}[f(\mathcal{I}_{t-1},\widehat{\theta}_n) - f(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]\sum_{t=n_0}^{n}\left(\frac{e^{i\langle s,K_t\rangle}-1}{N}\right)\right|_p^2 \\
&= 2I_1 + 2I_2. \qquad \text{(A.2)}
\end{aligned}
$$

First, we consider I_1. By a direction calculation,

$$
\begin{aligned}
I_1 = \sum_{j=1}^{p}\Bigg\{&\frac{1}{N^2}\left(\sum_{t=n_0}^{n}[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n) - f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)][\cos\langle s,K_t\rangle - 1]\right)^2 \\
&+ \frac{1}{N^2}\left(\sum_{t=n_0}^{n}[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n) - f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]\sin\langle s,K_t\rangle\right)^2\Bigg\}, \qquad \text{(A.3)}
\end{aligned}
$$

where f_j is the j-th component of f. Furthermore, we can show

$$
\begin{aligned}
&E\left(\sum_{t=n_0}^{n}[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n) - f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)][\cos\langle s,K_t\rangle - 1]\right)^2 \\
&\leq \left(\sum_{t=n_0}^{n}\left(E\left\{[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n) - f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)][\cos\langle s,K_t\rangle - 1]\right\}^2\right)^{1/2}\right)^2 \\
&\leq \left(\sum_{t=n_0}^{n}\left(E[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n) - f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]^4 E[\cos\langle s,K_t\rangle - 1]^4\right)^{1/4}\right)^2 \\
&\leq C\left(1 \wedge \|s\|^2\right)\left(\sum_{t=n_0}^{n}\left(E[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n) - f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]^4\right)^{1/4}\right)^2, \qquad \text{(A.4)}
\end{aligned}
$$

where the first inequality holds by Minkowski inequality, the second inequality by Hölder's inequality, and the third inequality holds by (A.1) and the fact that $[\cos\langle s,K_t\rangle - 1]^4 \leq 4[\cos\langle s,K_t\rangle - 1]^2$ and $E\|K_t\|^4 < \infty$ by Assumption 3.1(iii).

Similarly, we have

$$E\left(\sum_{t=n_0}^{n}[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n)-f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]\sin\langle s,K_t\rangle\right)^2$$

$$\leq C\left(1\wedge\|s\|^2\right)\left(\sum_{t=n_0}^{n}\left(E[f_j(\mathcal{I}_{t-1},\widehat{\theta}_n)-f_j(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]^4\right)^{1/4}\right)^2. \tag{A.5}$$

Hence, by (A.3)–(A.5), it follows that

$$EI_1\leq\frac{C\left(1\wedge\|s\|^2\right)}{N^2}\left(\sum_{t=n_0}^{n}\left(E\sup_{\theta\in\Theta}\|f(\mathcal{I}_{t-1},\theta)-f(\widehat{\mathcal{I}}_{t-1},\theta)\|^4\right)^{1/4}\right)^2. \tag{A.6}$$

Next, we consider I_2. By a direct calculation,

$$I_2=\left\|\frac{1}{N}\sum_{t=n_0}^{n}[f(\mathcal{I}_{t-1},\widehat{\theta}_n)-f(\widehat{\mathcal{I}}_{t-1},\widehat{\theta}_n)]\right\|^2\times\left|\sum_{t=n_0}^{n}\left(\frac{e^{i\langle s,K_t\rangle}-1}{N}\right)\right|^2.$$

Then, by Hölder's inequality, Minkowski inequality and a similar argument as for (A.4), we can prove

$$EI_2\leq\frac{C\left(1\wedge\|s\|^2\right)}{N^2}\left(\sum_{t=n_0}^{n}\left(E\sup_{\theta\in\Theta}\|f(\mathcal{I}_{t-1},\theta)-f(\widehat{\mathcal{I}}_{t-1},\theta)\|^4\right)^{1/4}\right)^2. \tag{A.7}$$

Now, (i) holds by (A.2), (A.6)–(A.7), and Assumption 3.5. This completes the proof. □

LEMMA A.3. *Suppose that $\widehat{\theta}_n-\theta_0=o_p(1)$, and Assumptions 3.1(i) and 3.2 hold. Then,*

$$\sup_{s\in\Omega}|\sqrt{N}\widetilde{\mathcal{G}}_n(s)-\sqrt{N}\mathcal{G}_n(s)+\Upsilon(s)[\sqrt{N}(\widehat{\theta}_n-\theta_0)]|=o_p(\sqrt{N}(\widehat{\theta}_n-\theta_0)),$$

on any compact set $\Omega\subset\mathcal{R}^k$, where $\Upsilon(s)$ is defined as in Theorem 3.1.

PROOF. By expanding $f(\cdot,\widehat{\theta}_n)$ at θ_0, we have

$$\widetilde{e}_t=e_t-g_t(\theta_0)(\widehat{\theta}_n-\theta_0)+[g_t(\theta_0)-g_t(\tilde{\theta}_n)](\widehat{\theta}_n-\theta_0), \tag{A.8}$$

where $\tilde{\theta}_n$ lies between θ_0 and $\widehat{\theta}_n$. Hence, it follows that

$$\sqrt{N}\widetilde{\mathcal{G}}_n(s)=\sqrt{N}\mathcal{G}_n(s)-\Upsilon_n(s)[\sqrt{N}(\widehat{\theta}_n-\theta_0)]+R_n(s),$$

where

$$\Upsilon_n(s)=\sum_{t=n_0}^{n}\frac{g_t(\theta_0)e^{i\langle s,K_t\rangle}}{N}-\sum_{t=n_0}^{n}\frac{g_t(\theta_0)}{N}\sum_{t=n_0}^{n}\frac{e^{i\langle s,K_t\rangle}}{N},$$

$$R_n(s) = \left\{ \sum_{t=n_0}^{n} \frac{[g_t(\theta_0) - g_t(\tilde{\theta}_n)]e^{i\langle s, K_t\rangle}}{N} - \sum_{t=n_0}^{n} \frac{(g_t(\theta_0) - g_t(\tilde{\theta}_n))}{N} \sum_{t=n_0}^{n} \frac{e^{i\langle s, K_t\rangle}}{N} \right\}$$
$$\times [\sqrt{N}(\widehat{\theta}_n - \theta_0)].$$

By Assumptions 3.1(i), 3.2 and the uniformly ergodic theorem (Jennich, 1969), $\Upsilon_n(s) = \Upsilon(s) + o_p(1)$. Moreover, since $\widehat{\theta}_n - \theta_0 = o_p(1)$, we can show that $R_n(s) = o_p(\sqrt{N}(\widehat{\theta}_n - \theta_0))$ by Assumptions 3.1(i) and 3.2, the uniformly ergodic theorem and the dominated convergence theorem. This completes the proof. □

LEMMA A.4. *Suppose that Assumptions 3.1–3.2 hold. Then,*

$$\sqrt{N}\mathcal{G}_{n,\varepsilon}(s) \Rightarrow \Delta(s),$$

on any compact set $\Omega \subset \mathcal{R}^k$, where $\Delta(s)$ is defined as in Theorem 3.1.

PROOF. Without loss of generality, we assume $N = n$. First, the finite-dimensional convergence of $\sqrt{n}\mathcal{G}_{n,\varepsilon}(s)$ follows directly from the martingale difference central limit theorem and the Cramér-Wold device.

Second, we show the tightness of $\sqrt{n}\mathcal{G}_{n,\varepsilon}(s)$ on compact sets by using the sufficient condition of Theorem 3 in Bickel and Wichura (1971) for multi-parameter processes. We evaluate the process on cubes $(s, v] = \Pi_{l=1}^{k}(s_l, v_l]$, where $s = (s_1, ..., s_k)^T$, $v = (v_1, ..., v_k)^T$, and $s_i < v_i$ for $i = 1, ..., k$. Recall that $K_t = (K_{1t}, ..., K_{kt})'$. Let

$$J_t^{u_1,\cdots,u_k}(s, v) = \prod_{l=1}^{k} e^{i(s_l + u_l(v_l - s_l))K_{lt}} - E\left[\prod_{l=1}^{k} e^{i(s_l + u_l(v_l - s_l))K_{l0}}\right],$$
$$\zeta_t(s, v) = \prod_{l=1}^{k}(e^{iv_l K_{lt}} - e^{is_l K_{lt}}) - E\left[\prod_{l=1}^{k}(e^{iv_l K_{l0}} - e^{is_l K_{l0}})\right].$$

Then, the increment of $\sqrt{n}\mathcal{G}_{n,\varepsilon}(s)$ on $(s, t]$ is given by

$$\begin{aligned}
&I_n(s, v] = \sqrt{n}\mathcal{G}_{n,\varepsilon}((s, v]) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \left\{ \sum_{u_1=0,1} \cdots \sum_{u_k=0,1} (-1)^{k - \sum_j u_j} J_t^{u_1,\cdots,u_k}(s, v) \right\} \\
&\quad - \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \right) \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{u_1=0,1} \cdots \sum_{u_k=0,1} (-1)^{k - \sum_j u_j} J_t^{u_1,\cdots,u_k}(s, v) \right) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \zeta_t(s, v) - \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \right) \left(\frac{1}{n} \sum_{t=1}^{n} \zeta_t(s, v) \right).
\end{aligned}$$

Below, we suppress the dependence of ζ_t on s and v. Let $W_t = \varepsilon_t \zeta_t$. By Cauchy-Schwarz inequality, it is straightforward to see

$$E|I_n(s,v)|_k^2 \le 2E\left[\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n W_t\right|_k^2\right] + 2E\left[\left|\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \varepsilon_t\right)\left(\frac{1}{n}\sum_{t=1}^n \zeta_t\right)\right|_k^2\right]$$

$$\equiv 2(I_3 + I_4). \tag{A.9}$$

For I_3, by the stationarity of W_t, we have

$$I_3 = E|W_0|_k^2 + 2\sum_{h=1}^{n-1}(1-h/n)\mathrm{Re}E[W_0^\star W_h] = E|W_0|_k^2, \tag{A.10}$$

where the second equality holds since W_t is an MDS with respect to $\mathcal{F}_{t-1}$. Next, we can show

$$E|W_0|_k^2 = E(\|\varepsilon_0\|^2|\zeta_0|^2) \le [E\|\varepsilon_0\|^4 E|\zeta_0|^4]^{1/2} \le C[E|\zeta_0|^4]^{1/2}, \tag{A.11}$$

where the first inequality holds by Hölder's inequality, and the second inequality is due to the fact that $E\|\varepsilon_0\|^4 < \infty$ by Assumption 3.1(ii). Note that there exists a constant $u_0 \in (1,2]$ such that $E\prod_{l=1}^k \|K_{l0}\|^{2u_0} < \infty$ by Assumption 3.1(iv), and then we have

$$\begin{aligned}
|\zeta_0|^2 &\le 2\left|\prod_{l=1}^k (e^{iv_lK_{l0}} - e^{is_lK_{l0}})\right|^2 + 2E\left|\prod_{l=1}^k (e^{iv_lK_{l0}} - e^{is_lK_{l0}})\right|^2 \\
&= 2\prod_{l=1}^k \left|e^{iv_lK_{l0}} - e^{is_lK_{l0}}\right|^2 + 2E\left[\prod_{l=1}^k \left|e^{iv_lK_{l0}} - e^{is_lK_{l0}}\right|^2\right] \\
&= 2\prod_{l=1}^k \left|e^{i(v_l-s_l)K_{l0}} - 1\right|^2 + 2E\left[\prod_{l=1}^k \left|e^{i(v_l-s_l)K_{l0}} - 1\right|^2\right] \\
&\le C\prod_{l=1}^k \left(1 \wedge (|v_l-s_l|\|K_{l0}\|)^2\right) + CE\left[\prod_{l=1}^k \left(1 \wedge (|v_l-s_l|\|K_{l0}\|)^2\right)\right] \\
&\le C\prod_{l=1}^k \left(1 \wedge (|v_l-s_l|\|K_{l0}\|)^{u_0}\right) + CE\left[\prod_{l=1}^k \left(1 \wedge (|v_l-s_l|\|K_{l0}\|)^{u_0}\right)\right] \\
&\le C\big(1 \wedge \prod_{l=1}^k (|v_l-s_l|\|K_{l0}\|)^{u_0}\big) + C\big(1 \wedge \prod_{l=1}^k |v_l-s_l|^{u_0}\big),
\end{aligned} \tag{A.12}$$

where the first inequality holds by Cauchy-Schwarz inequality, the second inequality holds by a similar argument as for (A.1), the third inequality holds since $u_0 \le 2$, and the fourth inequality holds since $E\prod_{l=1}^k \|K_{l0}\|^{2u_0} < \infty$. By (A.10)–(A.12) and Assumption 3.1(iv), it follows that

$$I_3 \le C\left(\prod_{l=1}^k |v_l-s_l|^{u_0}\right)\left(E\prod_{l=1}^k \|K_{l0}\|^{2u_0}\right)^{1/2} \le C\left(\prod_{l=1}^k |v_l-s_l|^{u_0}\right). \tag{A.13}$$

For I_4, by a similar argument as for (A.11), we have

$$I_4 \leq \left(n^2 E\left\|\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\right\|^4\right)^{1/2}\left(E\left|\frac{1}{n}\sum_{t=1}^{n}\zeta_t\right|^4\right)^{1/2}. \tag{A.14}$$

Note that

$$\begin{aligned} E\left|\frac{1}{n}\sum_{t=1}^{n}\zeta_t\right|^4 &\leq \frac{2}{n^4}\left[E\left(\sum_{t=1}^{n}\mathrm{Re}(\zeta_t)\right)^4 + E\left(\sum_{t=1}^{n}\mathrm{Im}(\zeta_t)\right)^4\right] \\ &\leq \frac{2}{n^4}\left[\sum_{t=1}^{n}[E\left(\mathrm{Re}(\zeta_t)^4\right)]^{1/4}\right]^4 + \frac{2}{n^4}\left[\sum_{t=1}^{n}[E\left(\mathrm{Im}(\zeta_t)^4\right)]^{1/4}\right]^4 \\ &\leq 2E\left|\zeta_t\right|^4 + 2E\left|\zeta_t\right|^4 = 4E\left|\zeta_t\right|^4, \end{aligned} \tag{A.15}$$

where the first inequality holds by Cauchy-Schwarz inequality, the second inequality holds by Minkowski inequality, and the third inequality holds by the stationarity of ζ_t and the fact that $\|\mathrm{Re}(\zeta_t)\| \leq |\zeta_t|$ and $\|\mathrm{Im}(\zeta_t)\| \leq |\zeta_t|$. Hence, by (A.12) and (A.15), we can show

$$E\left|\frac{1}{n}\sum_{t=1}^{n}\zeta_t\right|^4 \leq C\left[1 \wedge \prod_{l=1}^{k}|v_l - s_l|^{2u_0}\right], \tag{A.16}$$

where we have used the fact that $E\prod_{l=1}^{k}\|K_{l0}\|^{2u_0} < \infty$.

Let $\varepsilon_t = (\varepsilon_{t1}, ..., \varepsilon_{tp})^T$ and $S_{nl} = \sum_{t=1}^{n}\varepsilon_{lt}$ for $l = 1, ..., p$. Since ε_{lt} is an MDS with respect to $\mathcal{F}_{t-1}$, by Burkholder's inequality we have

$$\begin{aligned} n^2 E\left\|\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\right\|^4 &= \frac{1}{n^2}E\left(\sum_{l=1}^{p}S_{nl}^2\right)^2 \\ &\leq \frac{C}{n^2}\sum_{l=1}^{p}E\left(S_{nl}^4\right) \leq \frac{C}{n^2}\sum_{l=1}^{p}E\left(\sum_{t=1}^{n}\varepsilon_{lt}^2\right)^2 \leq C, \end{aligned} \tag{A.17}$$

and together with (A.14) and (A.16), it follows that

$$I_4 \leq C\left(\prod_{l=1}^{k}|v_l - s_l|^{u_0}\right). \tag{A.18}$$

Finally, by (A.9), (A.13), and (A.18), we can obtain

$$E|I_n(s,v]|_k^2 \leq C\left(\prod_{l=1}^{k}|v_l - s_l|^{u_0}\right).$$

Because $u_0 > 1$, the condition of Theorem 3 in Bickel and Wichura (1971) holds, and hence the processes $\sqrt{n}\mathcal{G}_{n,\varepsilon}(s)$ are tight on compact sets. This completes the proof. $\square$

Lemma A.5. *Suppose that Assumptions 3.1–3.4 hold. Then, under H_0,*

$$\sqrt{N}(\mathcal{G}_{n,\varepsilon}(s), \widehat{\theta}_n - \theta_0) \Rightarrow (\Delta(s), \mathcal{V}),$$

on any compact set $\Omega \subset \mathcal{R}^k$, where $\Delta(s)$ and $\mathcal{V}$ are defined as in Theorem 3.1.

Proof. By Assumption 3.4, the proof follows directly from Lemma A.4 and the martingale central limit theorem. □

Proof Theorem 3.1. Note that under H_0, $\mathcal{G}_n(s) = \mathcal{G}_{n,\varepsilon}(s)$ and $\sqrt{N}(\widehat{\theta}_n - \theta_0) = O_p(1)$ by Assumption 3.4. Then, by Lemmas A.2(ii) and A.3, we have

$$\sqrt{N}\widehat{\mathcal{G}}_n(s) = \sqrt{N}\mathcal{G}_{n,\varepsilon}(s) - \Upsilon(s)[\sqrt{N}(\widehat{\theta}_n - \theta_0)] + o_p(1),$$

on any compact set $\Omega \subset \mathcal{R}^k$. The conclusion follows by Lemma A.5 and the continuous mapping theorem. This completes the proof. □

Define $\mathcal{G}_{n,g}(s)$ in the same way as $\widehat{\mathcal{G}}_n(s)$ in (3.3) with $\widehat{e}_t$ replaced by $g_t(\tilde{\theta}_n)$, where $g_t(\tilde{\theta}_n)$ is defined as in (A.8). In order to prove Corollary 3.1, we need one more technical lemma.

Lemma A.6. *Suppose that Assumptions 3.1(i)-(iii) and 3.2 hold. Then,*
(i) $N \cdot E[|\mathcal{G}_{n,\varepsilon}(s)|_p^2] \leq C(1 \wedge \|s\|^2)$;
(ii) $E[|\mathcal{G}_{n,g}(s)|_p^2] \leq C(1 \wedge \|s\|^2)$,
for all $s \in \mathcal{R}^k$.

Proof. (i) Without loss of generality, we assume $N = n$. Let $\xi_t = e^{i\langle s,K_t\rangle} - E(e^{i\langle s,K_0\rangle})$ and $U_t = \varepsilon_t\xi_t$, where we suppress the dependence of ξ_t (or U_t) on s. Then, as for (A.9), we have

$$\begin{aligned} nE|\mathcal{G}_{n,\varepsilon}(s)|_p^2 &\leq 2E\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\right|_p^2 + 2E\left|\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\right)\left(\frac{1}{n}\sum_{t=1}^{n}\xi_t\right)\right|_p^2 \\ &\equiv 2(I_5 + I_6). \end{aligned} \tag{A.19}$$

By a similar argument as for (A.1) and (A.10)–(A.11), we can show

$$I_5 \leq C[E|\xi_0|^4]^{1/2} \leq C(1 \wedge \|s\|^2), \tag{A.20}$$

where we have used the fact that $E\|\varepsilon_0\|^4 < \infty$ by Assumption 3.1(ii) and $E\|K_0\|^4 < \infty$ by Assumption 3.1(iii).

Moreover, by (A.20), a similar argument as for (A.14)–(A.15) entails

$$I_6 \leq C \left(n^2 E \left\| \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \right\|^4 \right)^{1/2} \left(E \left| \frac{1}{n} \sum_{t=1}^{n} \xi_t \right|^4 \right)^{1/2}, \tag{A.21}$$

where

$$E \left| \frac{1}{n} \sum_{t=1}^{n} \xi_t \right|^4 \leq 4E \left|\xi_t\right|^4 \leq C(1 \wedge \|s\|^4).$$

By (A.17) and (A.21), it follows that

$$I_6 \leq C(1 \wedge \|s\|^2). \tag{A.22}$$

Hence, (i) holds by (A.19)–(A.20) and (A.22).

(ii) As for (A.2),

$$\begin{aligned} |\mathcal{G}_{n,g}(s)|_p^2 &\leq 2 \left| \frac{1}{n} \sum_{t=1}^{n} g_t(\tilde{\theta}_n)(e^{i\langle s, K_t\rangle} - 1) \right|_p^2 \\ &\quad + 2 \left| \left(\sum_{t=1}^{n} \frac{g_t(\tilde{\theta}_n)}{n} \right) \left(\sum_{t=1}^{n} \frac{e^{i\langle s, K_t\rangle} - 1}{n} \right) \right|_p^2 \\ &= 2I_7 + 2I_8. \end{aligned} \tag{A.23}$$

By the similar arguments as for (A.6)–(A.7), it is not hard to see

$$\begin{aligned} EI_7 &\leq \frac{C\left(1 \wedge \|s\|^2\right)}{n^2} \left(\sum_{t=1}^{n} \left(E \sup_{\theta \in \Theta} \|g_t(\theta)\|_F^4 \right)^{1/4} \right)^2 \leq C\left(1 \wedge \|s\|^2\right), \\ EI_8 &\leq \frac{C\left(1 \wedge \|s\|^2\right)}{n^2} \left(\sum_{t=1}^{n} \left(E \sup_{\theta \in \Theta} \|g_t(\theta)\|_F^4 \right)^{1/4} \right)^2 \leq C\left(1 \wedge \|s\|^2\right), \end{aligned}$$

where we have used Assumption 3.2 and the stationarity of $g_t(\theta)$. By (A.23), it follows that (ii) holds. This completes all of the proofs. □

Proof of Corollary 3.1. Consider the compact set

$$\Omega_\rho = \{s \in \mathcal{R}^k : \rho \leq \|s\| \leq 1/\rho\} \quad \text{for small } \rho > 0. \tag{A.24}$$

By Theorem 3.1 and the continuous mapping theorem, we have under H_0,

$$N \left\| \frac{1}{c_k} \int_{\Omega_\rho} \frac{\widehat{\mathcal{G}}_n(s) \widehat{\mathcal{G}}_n(s)^\star}{\|s\|^{1+k}} ds \right\|_F \to_d \left\| \frac{1}{c_k} \int_{\Omega_\rho} \frac{\chi(s)\chi(s)^\star}{\|s\|^{1+k}} ds \right\|_F.$$

It remains to check that under H_0,

$$\lim_{\rho\downarrow 0}\limsup_{n\to\infty}\frac{N}{c_k}\int_{\Omega_\rho^c}\frac{\widehat{\mathcal{G}}_n(s)\widehat{\mathcal{G}}_n(s)^\star}{\|s\|^{1+k}}ds=0, \tag{A.25}$$

in probability. Note that under H_0, $\mathcal{G}_n(s)=\mathcal{G}_{n,\varepsilon}(s)$, and hence by (A.8),

$$\sqrt{N}\widehat{\mathcal{G}}_n(s)=\sqrt{N}[\widehat{\mathcal{G}}_n(s)-\widetilde{\mathcal{G}}_n(s)]+\sqrt{N}\mathcal{G}_{n,\varepsilon}(s)-\mathcal{G}_{n,g}(s)[\sqrt{N}(\widehat{\theta}_n-\theta_0)].$$

By Lemma A.6 and Markov's inequality, it is not hard to see

$$\lim_{\rho\downarrow 0}\limsup_{n\to\infty}\frac{N}{c_k}\int_{\Omega_\rho^c}\frac{\mathcal{G}_{n,\varepsilon}(s)\mathcal{G}_{n,\varepsilon}(s)^\star}{\|s\|^{1+k}}ds=0, \tag{A.26}$$

$$\lim_{\rho\downarrow 0}\limsup_{n\to\infty}\frac{1}{c_k}\int_{\Omega_\rho^c}\frac{\mathcal{G}_{n,g}(s)\mathcal{G}_{n,g}(s)^\star}{\|s\|^{1+k}}ds=0, \tag{A.27}$$

in probability, and a similar result holds for the cross term between $\sqrt{N}\mathcal{G}_{n,\varepsilon}(s)$ and $\mathcal{G}_{n,g}(s)$ by Hölder's inequality. Hence, the result (A.25) follows by Lemma A.2(i) and the fact that $\sqrt{N}(\widehat{\theta}_n-\theta_0)=O_p(1)$. This completes the proof. □

Proof of Theorem 3.2. Under H_a,

$$\begin{aligned}\widehat{e}_t&=[\widehat{e}_t-\widetilde{e}_t]+[f(\mathcal{I}_{t-1},\theta_*)-f(\mathcal{I}_{t-1},\widehat{\theta}_n)]+a_t+\varepsilon_t\\&=[\widehat{e}_t-\widetilde{e}_t]-g_t(\widetilde{\theta}_{n*})(\widehat{\theta}_n-\theta_*)+a_t+\varepsilon_t,\end{aligned} \tag{A.28}$$

where $\widetilde{\theta}_{n*}$ lies between θ_* and $\widehat{\theta}_n$.

Define $\mathcal{G}_{n,g_*}(s)$ and $\mathcal{G}_{n,a}(s)$ in the same way as $\widehat{\mathcal{G}}_n(s)$ in (3.3) with $\widehat{e}_t$ replaced by $g_t(\widetilde{\theta}_{n*})$ and a_t, respectively. Then, by (A.28), it follows that

$$\widehat{\mathcal{G}}_n(s)=[\widehat{\mathcal{G}}_n(s)-\widetilde{\mathcal{G}}_n(s)]-\mathcal{G}_{n,g_*}(s)(\widehat{\theta}_n-\theta_*)+\mathcal{G}_{n,a}(s)+\mathcal{G}_{n,\varepsilon}(s).$$

On any compact set $\Omega\subset\mathcal{R}^k$,

$$\widehat{\mathcal{G}}_n(s)-\widetilde{\mathcal{G}}_n(s)=o_p(1), \tag{A.29}$$

$$\mathcal{G}_{n,g_*}(s)=O_p(1), \tag{A.30}$$

$$\mathcal{G}_{n,a}(s)=\mathcal{A}(s)+o_p(1)\quad\text{and}\quad\mathcal{G}_{n,\varepsilon}(s)=o_p(1), \tag{A.31}$$

where (A.29) holds by Lemma A.2(ii), (A.30) holds by the uniformly ergodic theorem and the dominated convergence theorem, and (A.31) holds by the uniformly ergodic theorem. Since $\widehat{\theta}_n-\theta_*=o_p(1)$ by Assumption 3.3, the conclusion follows directly by (A.29)–(A.31). This completes the proof. □

Proof Theorem 3.3. Under $H_{a,n}$, by (A.8), we have

$$\sqrt{N}\widehat{\mathcal{G}}_n(s) = \sqrt{N}[\widehat{\mathcal{G}}_n(s) - \widetilde{\mathcal{G}}_n(s)] - \mathcal{G}_{n,g}(s)[\sqrt{N}(\widehat{\theta}_n - \theta_0)] + \mathcal{G}_{n,a}(s) + \sqrt{N}\mathcal{G}_{n,\varepsilon}(s).$$

Similar to Lemma A.5, we can show that under $H_{a,n}$,

$$\sqrt{N}(\mathcal{G}_{n,\varepsilon}(s), \widehat{\theta}_n - \theta_0) \Rightarrow (\Delta(s), \xi_a + \mathcal{V}), \tag{A.32}$$

on any compact set $\Omega \subset \mathcal{R}^k$. Since $\mathcal{G}_{n,g}(s) = \Upsilon(s) + o_p(1)$ and $\mathcal{G}_{n,a}(s) = \mathcal{A}(s) + o_p(1)$ uniformly on any compact set $\Omega \subset \mathcal{R}^k$, the conclusion holds by Lemma A.2(ii), (A.32) and the continuous mapping theorem. This completes the proof. □

Proof Corollary 3.2. By Theorem 3.3 and the continuous mapping theorem, it follows that under $H_{a,n}$,

$$N\left\|\frac{1}{c_k}\int_{\Omega_\rho}\frac{\widehat{\mathcal{G}}_n(s)\widehat{\mathcal{G}}_n(s)^\star}{\|s\|^{1+k}}ds\right\|_F \to_d \left\|\frac{1}{c_k}\int_{\Omega_\rho}\frac{\chi_a(s)\chi_a(s)^\star}{\|s\|^{1+k}}ds\right\|_F,$$

where Ω_ρ is defined in (A.24).

It remains to check (A.25) under $H_{a,n}$. Since $E\|a_t\|^4 < \infty$, by a similar argument as for Lemma A.6(ii), we can show that $E[|\mathcal{G}_{n,a}(s)|_p^2] \le C(1 \wedge \|s\|^2)$, and hence

$$\lim_{\rho\downarrow 0}\limsup_{n\to\infty}\frac{1}{c_k}\int_{\Omega_\rho^c}\frac{\mathcal{G}_{n,a}(s)\mathcal{G}_{n,a}(s)^\star}{\|s\|^{1+k}}ds = 0, \tag{A.33}$$

in probability. By (A.26)–(A.27) and (A.33), we can prove (A.25) under $H_{a,n}$. This completes the proof. □

Proof of Theorem 4.1. Without loss of generality, we assume $N = n$. Let $\delta_t^* = f(\mathcal{I}_{t-1}, \widehat{\theta}_n^*) - f(\widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n^*)$ and $\delta_t = f(\mathcal{I}_{t-1}, \widehat{\theta}_n) - f(\widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)$. By construction,

$$\begin{aligned}\widehat{e}_t^{**} &= [\delta_t^* - \delta_t] - [f(\mathcal{I}_{t-1}, \widehat{\theta}_n^*) - f(\mathcal{I}_{t-1}, \widehat{\theta}_n)] + \widehat{e}_t w_t^* \\ &= [\delta_t^* - \delta_t] - g_t(\widetilde{\theta}_n^*)(\widehat{\theta}_n^* - \widehat{\theta}_n) + \widehat{e}_t w_t^*,\end{aligned} \tag{A.34}$$

where $\widetilde{\theta}_n^*$ lies between $\widehat{\theta}_n^*$ and $\widehat{\theta}_n$. Then, it follows that

$$\begin{aligned}\sqrt{n}\widehat{\mathcal{G}}_n^*(s) &= [\sqrt{n}\widehat{\mathcal{G}}_{n,\delta}^*(s) - \sqrt{n}\widehat{\mathcal{G}}_{n,\delta}(s)] \\ &\qquad - \widehat{\mathcal{G}}_{n,g}^*(s)[\sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n)] + \sqrt{n}\widehat{\mathcal{G}}_{n,w}^*(s),\end{aligned} \tag{A.35}$$

where $\widehat{\mathcal{G}}_{n,\delta}^*(s)$, $\widehat{\mathcal{G}}_{n,\delta}(s)$, $\widehat{\mathcal{G}}_{n,g}^*(s)$, and $\widehat{\mathcal{G}}_{n,w}^*(s)$ are defined in the same way as $\widehat{\mathcal{G}}_n(s)$ in (3.3) with $\widehat{e}_t$ replaced by δ_t^*, δ_t, $g_t(\widetilde{\theta}_n^*)$, and $\widehat{e}_t w_t^*$, respectively.

By a similar argument as for Lemmas A.2–A.3, we can show that uniformly on any compact set $\Omega \subset \mathcal{R}^k$, $\sqrt{n}\widehat{\mathcal{G}}^*_{n,\delta}(s) = o_p^*(1)$, $\sqrt{n}\widehat{\mathcal{G}}_{n,\delta}(s) = o_p(1)$, and $\widehat{\mathcal{G}}^*_{n,g}(s) = \Upsilon_*(s) + o_p^*(1)$, and hence by (A.35),

$$\text{(A.36)} \qquad \sqrt{n}\widehat{\mathcal{G}}^*_n(s) = \sqrt{n}\widehat{\mathcal{G}}^*_{n,w}(s) - \Upsilon_*(s)[\sqrt{n}(\widehat{\theta}^*_n - \widehat{\theta}_n)] + o_p^*(1).$$

Furthermore, by Assumption 4.1 and a similar argument as for Lemmas A.4–A.5, we can prove that conditional on $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$,

$$\text{(A.37)} \qquad \sqrt{N}(\widehat{\mathcal{G}}^*_{n,w}(s), \widehat{\theta}^*_n - \widehat{\theta}_n) \Rightarrow (\Delta_*(s), \mathcal{V}_*) \text{ in probability,}$$

on any compact set $\Omega \subset \mathcal{R}^k$. Now, the conclusion follows by (A.36)–(A.37). This completes the proof. □

Proof of Corollary 4.1. By a similar argument as for Lemmas A.2 and A.6, we can show that $N \cdot E[|\widehat{\mathcal{G}}_{n,\delta}(s)|_p^2] \le C(1 \wedge \|s\|^2)$, and conditional on $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$, $N \cdot E^*[|\widehat{\mathcal{G}}^*_{n,\delta}(s)|_p^2] \le C(1 \wedge \|s\|^2)$, $E^*[|\widehat{\mathcal{G}}^*_{n,g}(s)|_p^2] \le C(1 \wedge \|s\|^2)$, and $N \cdot E^*[|\widehat{\mathcal{G}}^*_{n,w}(s)|_p^2] \le C(1 \wedge \|s\|^2)$ in probability. Then, the conclusion holds by (A.35) and a similar argument as for Corollary 3.1. This completes the proof. □

REFERENCES

[1] Bickel, P.J. and Wichura, M.J. (1971) Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Statistics* **42**, 1656-1670.

[2] Escanciano J.C. (2006) Goodness-of-fit tests for linear and non-linear time series models. *Journal of the American Statistical Association* **101**, 531-541.

[3] Escanciano, J.C. and Velasco, C. (2006) Generalized spectral tests for the martingale difference hypothesis. *Journal of Econometrics* **134**, 151-185.

[4] Jennrich, R.I. (1969) Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* **40**, 633-643.

[5] O'Hara M., Yao, C. and Ye, M. (2014) What's not there: Odd lots and market data. *Journal of Finance* **69**, 2199-2236.

[6] Székely, G.J. and Rizzo, M.L. (2005) Hierarchical clustering via joint between-within distance: extending wards minimum variance method. *Journal of Classification* **22**, 151-183.