Statistica Sinica Preprint No: SS-2024-0035	
Title	An Automatic MDDM-Based Test for Martingale Difference Hypothesis
Manuscript ID	SS-2024-0035
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0035
Complete List of Authors	Chenglong Zhong and Guochang Wang
Corresponding Authors	Guochang Wang
E-mails	wanggc023@amss.ac.cn
Notice: Accepted author version.	

An automatic MDDM-based test for martingale difference hypothesis

Chenglong Zhong, Guochang Wang*

School of Economics, Jinan University, Guangzhou, China.

Abstract: Checking whether the error term is a marginal difference sequence (MDS) in the multivariate time series model with a parametric conditional mean is a crucial problem. Tests based on the martingale difference divergence matrix (MDDM) are an effective statistical method for testing MDS in the residuals of multivariate time series models. However, MDDM-based tests require specifying the lag order. To solve this problem, we propose a data-driven MDDM-based test that automatically selects the lag order. This method has three main advantages: first, researchers do not need to specify the lag order while the test automatically selects it from the data; second, under the null hypothesis, the lag order is one; third, the proposed automatic tests have good performance in detecting model inadequacy caused by high-order dependence. In theory, we prove the asymptotical property of the proposed method. Furthermore, we demonstrate the effectiveness of this method through simulations and real data analysis.

Key words and phrases: Martingale difference divergence matrix (MDDM), Marginal difference sequence (MDS), Multivariate time series model.

*Corresponding author.

1. Introduction

The concept of marginal difference sequence (MDS) is central in many areas of economics and finance. Many economic theories in a dynamic context, such as market efficiency hypothesis, rational expectations, or optimal consumption smoothing, deliver such dependence restrictions on the underlying economic variables, e.g., Hall (1978) and Lo (1997). The martingale difference hypothesis (MDH) states that the best predictor, in the sense of least mean squared error, of the future values of a time series given the current information set is just the unconditional expectation. Hence, past information does not help to improve the forecast of future values of an MDS. Furthermore, MDH can be used to test the error term with parametric conditional mean for the time series model. More formally, people consider a 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 an MDS with respect to $\mathcal{I}_{t-1}$; and the conditioning set at time t is $\mathcal{I}_t = \{Y_t, Y_{t-1}, \cdots\}$. In general, empirical researchers usually assume that $m(\cdot)$ is from a parametric model, i.e., $m(\cdot) \in \mathcal{M}$, 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 $\Theta \in \mathcal{R}^s$. Correctly specifying the form of $m(\cdot)$ is crucial because the inference of θ and the prediction of future values of Y_t heavily rely on the form of $m(\cdot)$. To check whether the form of $m(\cdot)$ is correctly specified or not,

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 an MDH:

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

where $e_t = e_t(\theta_0)$ with $e_t(\theta) = Y_t - f(\mathcal{I}_{t-1}, \theta)$ for some $\theta_0 \in \Theta$. If H_0 is true, then the best predictor, in the sense of the 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, then 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, valid testing tools for H_0 must be provided.

In recent years, two different classes of MDH testing methods have been developed: the first one is to test the MDH for the observed data Y_t, i.e., under the assumption that $f(\mathcal{I}_{t-1}, \theta_0) = \mu$ (a constant scalar), these methods include but are not limited to Durlauf (1991), who proposed a MDH method for the univariate time series Y_t. Durlauf (1991), Hong (1996), Deo (2000), and Shao (2011a) investigated spectral tests, which target nonzero serial correlations. The second class of MDH testing methods is to test the MDH for the error term of the parametric conditional mean model, i.e., under the assumption that $f(\mathcal{I}_{t-1}, \theta_0) \not\equiv \mu$. Hong and Lee (2005) proposed the kernel-based spectral

test, and Escanciano (2006) proposed the generalized spectral test. For other spectral tests on serial dependence hypotheses or for model checks, one can refer to Hong (1996), Paparoditis (2000), Delgado et al. (2005), Shao (2011b), Chen and Hong (2014), Zhu and Li (2015), and many others. All the above methods are useful, but these methods are designed for $p = 1$. In the case of $p > 1$, Hosking (1980) and Li and McLeod (1981) proposed Portmanteau tests based on residual autocorrelations. However, these residual-autocorrelation-based tests are expected to exhibit a lack of power to the non-MDS e_t with zero autocorrelations, and their generalizations to other nonlinear multivariate models are not straightforward. To overcome the disadvantages of portmanteau tests, Wang et al. (2022) proposed a test for the multivariate MDH in (1.2) on the basis of the martingale difference divergence matrix (MDDM). 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. In contrast to the residual-autocorrelation-based tests that can only detect linear dependents, MDDM-based tests can detect linear and nonlinear dependents.

For most of the aforementioned methods, the lag order must be specified, and the selection of the employed number of lag M is arbitrary. Regarding this limitation, Wang et al.(2022) proposed computing the MDDM-based statistic for various lags M. However, as presented in the real data analysis, different lags lead to conflicting results. Selecting a lag order M in prior without considering the structure of the data may also lead to another problem in which the selected lag cannot accurately reflect the true structure of the data; furthermore, this affects the accuracy and reliability of the model. For an example, if the

data contain a high-order dependent, the tests may suffer inefficiency when the selected lag M does not contain the high-order dependent, i.e., the selected lag order is smaller than the true lag of the dependent.

To deal with this problem, we propose an MDDM-based statistic, which can select the lag of M automatically adapting to the lag of the dependent present in the data. Thus, under the null hypothesis (1.2), the lag M is one. However, under the alternative, the test would employ a higher value of M according to the serial dependent in the data. The selection of lag M is similar to the selection of the number of autocorrelations in the framework of testing for serial correlation. Thus, following Escanciano and Lobato (2009), which combined the advantages of AIC (Akaike,1974) and BIC (Schwarz,1978). On the one hand, tests constructed using the BIC criterion can effectively control the type-I error and are more powerful when the dependent is present in the first lag dependence. On the other hand, tests based on AIC cannot effectively control the type-I error, but they are more powerful when the dependent is present in the high-order dependence. As shown in the simulation study of Subsection 5.3, our statistic has good performance in detecting the model inadequacy caused by high-order dependence, while for the MDDM-based testing methods, the power is very low especially for the small value of lag order M. The detailed explanation and introduction for the data-driven MDDM-based methods are presented in Section 3.

The rest of the article is organized as follows: Section 2 reviews the MDDM-based test statistic. Section 3 presents our proposed tests and establishes its asymptotic properties. Section 4 provides a bootstrap method for estimating the critical values of our proposed

tests. Section 5 studies the finite sample behavior through simulations. Section 6 makes an empirical analysis of our proposed test statistics. Conclusion and discussion are offered in Section 7. Proofs, the simulation studies including for $p = 1$ and the moment assumptions failed are gathered in the Supplementary Materials.

Throughout the paper, $\mathcal{R}$ is a one-dimensional real vector space, $\mathcal{C}$ is a 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 $x \in \mathcal{R}^p$, and $\langle x, y\rangle$ is the inner product for $x, y \in \mathcal{R}^p$. For a matrix $X \in \mathcal{R}^{p\times q}$, $X^\top$ is its transpose, $\|X\|_F$ is its Frobenius norm, and $\|X\|_2$ is its spectral norm, $\mathrm{vec}(X)$ is its vectorization. Moreover, 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. Preliminaries

In this section, we first introduce the MDDM (Lee and Shao, 2018), which forms the MDDM-based 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 $\mathcal{G}(s) = \mathrm{cov}(V, e^{i\langle s,U\rangle}) = (\mathcal{G}_1(s), \cdots, \mathcal{G}_p(s))^\top$ 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 $E[\|U\|^2 + \|V\|^2] < \infty$, as shown in Lee and Shao (2018),

then the MDDM has a simple and equivalent expression as follows:

$$\text{MDDM}(V|U) = -E[(V - E(V))(V' - E(V'))^\top \|U - U'\|], \tag{2.1}$$

where (V', U') is an independent and identically distributed (i.i.d.) copy of (V, U). For the sake of completeness, we refer the readers to Lee and Shao (2018) and Wang et al. (2022) for the proof of this equivalence.

Let random sample $\{(U_k, V_k)\}_{k=1}^n$ from the joint distribution of (U, V), and the result (2.1) implies that the sample MDDM (denoted by $\text{MDDM}_n(V|U)$) can be computed as

$$\text{MDDM}_n(V|U) = -\frac{1}{n^2}\sum_{h,l=1}^{n}(V_h - \overline{V}_n)(V_l - \overline{V}_n)^\top \|U_h - U_l\|,$$

where $\overline{V}_n = n^{-1}\sum_{i=1}^{n} V_i$ (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:

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

Hence, we can test the hypothesis of $E(V|U) = 0$ by examining whether $\|\text{MDDM}_n(V|U)\|_F$ is significantly different from zero. In view of this fact, Wang et al.(2022) proposed two MDDM-based test statistic $\widehat{T}_{sn}^F(M)$ and $\widehat{T}_{wn}^F(M)$ for H_0 as follows:

$$\widehat{T}_{sn}^F(M) = n\|\text{MDDM}_n(\widehat{e}_t|Y_{t,1:M})\|_F, \tag{2.2}$$

and

$$\widehat{T}_{wn}^F(M) = n\sum_{j=1}^{M}\omega_j\|\text{MDDM}_n(\widehat{e}_t|Y_{t-j})\|_F, \tag{2.3}$$

where $Y_{t,1:M} := (Y_{t-1}^\top, \cdots, Y_{t-M}^\top)^\top$, $\omega_j = (n-j)/n$, $\widehat{e}_t = Y_t - f(\widehat{\mathcal{I}}_t, \widehat{\theta}_n)$ is the residual, $\widehat{\mathcal{I}}_t$ is the observed conditional set at time t, and $\widehat{\theta}_n$ is an estimator of θ_0. To study whether the MDDM-based test is influenced by the matrix norm, Wang et al. (2022) also suggested to replace the Frobenius norm with the spectral norm and proposed two another MDDM-based tests denoted by $\widehat{T}_{sn}^S(M)$ and $\widehat{T}_{wn}^S(M)$. Thus, there are four MDDM-based test statistics, $\widehat{T}_{sn}^F(M)$, $\widehat{T}_{wn}^F(M)$, $\widehat{T}_{sn}^S(M)$, and $\widehat{T}_{wn}^S(M)$ proposed by Wang et al.(2022).

The MDDM-based tests have appealing theoretical properties and good numerical performance, but the MDDM-based tests should give a fixed lag M in advance, and Wang et al. (2022) proposed to avoid this disadvantage by considering a variety of M. However, the potential issue of conflicting evidence among different lag orders M has been ignored. This conflict may lead to the selected lag order not accurately reflecting the true structure of the data, thereby affecting the accuracy and reliability of the model. Additionally, when the selected lag order M is smaller than the true lag dependent in the data, the tests lose some power because the high-order dependence is not considered. A simulation studies of a high-order dependent is designed in Subsection 5.3, which shows that the MDDM-based tests proposed by Wang et al. (2022) is inefficient in the high-order dependence, especially for a small value of M. To address this issue, we develop a data-driven method for automatically selecting the lag order M, as detailed in Section 3. From the numerical study of Wang et al. (2022), the weighted MDDM-based tests $\widehat{T}_{wn}^F(M)$ and $\widehat{T}_{wn}^S(M)$ are usually more powerful than the tests $\widehat{T}_{sn}^F(M)$ and $\widehat{T}_{sn}^S(M)$. This result indicates that leveraging weighted MDDM-based tests could potentially enhance the sensitivity of our tests when choosing lag orders. Alternatively, the automatic methods

proposed by Inglot and Ledwina (2006a, b) and Escanciano and Lobato (2009) need the test statistic increasing with the lag of M increasing, and the tests for $\widehat{T}_{sn}^{F}(M)$ and $\widehat{T}_{sn}^{S}(M)$ do not necessarily increase as M increases. Thus, we do not consider the lag in the selected method for the tests $\widehat{T}_{sn}^{F}(M)$ and $\widehat{T}_{sn}^{S}(M)$.

3. Data-driven MDDM-based test

3.1 Data-driven test statistics

We propose to modify the statistic (2.3) by allowing the data to select the lag M automatically. In particular, the proposed test statistic is the maximized value of statistic (2.3) corrected by a penalty term that is an increasing function of the included number of lag M. Furthermore, we allow the data to select whether AIC or BIC is employed as the penalty function. The motivation for letting the data select AIC or BIC is as follows. On the one hand, the AIC criterion imposes a small penalty that naturally leads to large values for the selected M. However, Akaike's small penalty implies that the AIC criterion is inconsistent in Woodroofe (1982); thus, under the null hypothesis, it leads to a test that cannot control the type-I error. On the other hand, the large penalty imposed by the BIC criterion implies that the chosen values for M are low. Small values for M lead to tests that control the type-I error properly, but are less powerful when the dependence appears at higher lags. To obtain the best from both worlds, the ideal test procedure employs the BIC when the evidence points to the null hypothesis (the sample MDDMs appear to be small) and employs the AIC when the evidence points to the alternative (some sample MDDM appears to be large).

Formally, the proposed test statistic takes the form

$$\widehat{AT}_{wn}^{F} = \widehat{T}_{wn}^{F}(M^*),$$

where

$$M^* = \min\{M : 1 \le M \le d; L_M \ge L_h, h = 1, 2, \cdots, d\},$$

$$L_M = n\|\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-M})\|_F - \log(3p)\pi(M, n, k),$$

d is a fixed upper bound, and $\pi(M, n, k)$ is a penalty term that takes the form

$$\pi(M, n, k) = \begin{cases} M \log n, & \text{if } \max\limits_{1\le j\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F \le \log(3p)^k\sqrt{\log n}, \\ 2M, & \text{if } \max\limits_{1\le j\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F > \log(3p)^k\sqrt{\log n}, \end{cases} \tag{3.1}$$

where k is some fixed positive number and calls $\log(3p)^k\sqrt{\log n}$ as the penalty term. We denote the data-driven MDDM-based statistic as $\widehat{AT}_{wn}^{F}$. We also can obtain $\widehat{AT}_{wn}^{S}$ just by replacing the Frobenius norm with the spectral norm. Then, we propose two data-driven MDDM-based statistics denoted by $\widehat{AT}_{wn}^{F}$ and $\widehat{AT}_{wn}^{S}$, respectively. For simplicity, we consider d to be a large fixed positive integer number. At the end of this section, we will give a comment about d and consider that d grows slowly to infinity with the sample size n. To implement the automatic selection procedure, we need to decide the other turning parametric k. A small value of k could lead to the choice of the AIC criterion, while a large k would lead to the choice of the BIC criterion. Moderate values provide a "switching effect" in which one combines the advantages of the two selection rules. From Subsection 5.1, we know that $k = 1.8$ is a reasonable value in all of the numerical studies. Following, we will give a remark about the selected penalty term.

Remark 1. Too large a value of the penalty leads to the data-driven MDDM-based test always choosing the BIC and resulting in low power under the alternative hypothesis. Moreover, the test will always select the AIC if the value of the penalty is too small; thus, the data-driven MDDM-based test cannot control the type-I error. Therefore, a suitable choice of the penalty term is essential. The motivation of considering the penalty term is as follows: First, compared with the automatic Portmanteau test of Escanciano and Lobato (2009), the MDDM-based test has the same convergence rate as the Portmanteau test; then, we can use $\sqrt{\log n}$ in the penalty term. Second, given that the automatic Portmanteau test of Escanciano and Lobato (2009) only considers the scalar time series, the penalty term does not contain the dimensionality of the time series p. By contrast, our data-driven MDDM-based test is designed for multivariate series; thus, our proposed data-driven MDDM-based test should contain p. By combining with the theoretical result of Theorem 1 and that of vast simulation studies, we find that the term $\log(3p)^k$ is a suitable choice, where k is some fixed positive number. Thus, we choose $\log(3p)^k\sqrt{\log n}$ as the penalty term. The numerical studies show that the selected penalty term is a suitable one.

3.2 Theoretical properties of data-driven MDDM-based test

In this subsection, we study the asymptotics of $\widehat{AT}_{wn}^{F}$. Similar results can be shown for $\widehat{AT}_{wn}^{S}$ with a slight modification, and details are omitted. We first give some regularity conditions. 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

$Y_t = (Y_{1,t}, \cdots, Y_{p,t})^\top$.

Assumption 1. *(i)* $\{(Y_t, \varepsilon_t)\}$ *is strictly stationary and ergodic; (ii)* $E\|\varepsilon_t\|^4 < \infty$*; (iii)* $E\|Y_t\|^4 < \infty$*; (iv)* $E\prod_{k=1}^{p}\|Y_{k,t}\|^{2u} < \infty$ *for some* $u \in (1, 2]$.

Assumption 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. *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 4. *The estimator* $\widehat{\theta}_n$ *satisfies the asymptotic expansion under* H_0*,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Upsilon(Y_t, \mathcal{I}_{t-1}, \theta_0) + o_p(1), \tag{3.2}$$

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

Assumption 5. *For* $\widehat{\mathcal{I}}_t$ *given in* $\widehat{e}_t = Y_t - f(\widehat{\mathcal{I}}_t, \widehat{\theta}_n)$*,*

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

Assumptions 1 – 5 are the same as those in Wang et al.(2022), which are used to ensure that the MDDM-based test statistics converge to a random processor under the null hypothesis and the MDDM-based test is consistent under the alternative hypothesis. The detailed explanation for these assumptions can be found in Wang et al.(2022).

To elaborate the asymptotic distribution of $\widehat{AT}_{wn}^F$, we need more notation. Let $\phi^j(s) = \mathrm{cov}(g_t(\theta_0), e^{i\langle s, Y_{t-j}\rangle})$, $\mathcal{V}$ be a normal random vector with mean zero and variance-covariance

matrix $L(\theta_0)$, $\Delta^j(s)$ be a complex valued Gaussian field with mean zero and covariance matrix function $\text{cov}(\Delta^j(s), \Delta^j(s')^\star) = E(\varepsilon_t \varepsilon_t^\top e^{i[\langle s, Y_{t-j}\rangle + \langle s', Y_{t-j}\rangle]})$, and $(\Delta^j(s), \mathcal{V})$ be jointly Gaussian with covariance matrix $\text{cov}(\Delta(s), \mathcal{V}) = E(\varepsilon_t \Upsilon(Y_t, \mathcal{I}_{t-1}, \theta_0)^\top e^{i\langle s, Y_{t-j}\rangle})$.

First, we study the asymptotic of $\widehat{AT}_{wn}^F$ under the null hypothesis. On the basis of Theorem 3.1 in Wang et al. (2022), we can establish the asymptotic distribution in the following lemma:

Lemma 1. *Suppose that Assumptions 1-5 hold. Then, under* H_0,

$$(n-j)\|\text{MDDM}_n(\widehat{e}_t|Y_{t-j})\|_F \to_d \left\| \frac{1}{c_p} \int_{\mathcal{R}^p} \frac{\chi^j(s)\chi^j(s)^\star}{\|s\|^{1+p}} ds \right\|_F,$$

where $\chi^j(s) = \Delta^j(s) - \phi^j(s)\mathcal{V}$.

On the basis of Lemma 1, we can establish the asymptotic null distribution of $\widehat{AT}_{wn}^F$ in the following theorem:

Theorem 1. *Suppose that Assumptions 1–5 hold. Then, under* H_0,

$$\widehat{AT}_{wn}^F \to_d \left\| \frac{1}{c_p} \int_{\mathcal{R}^p} \frac{\chi^1(s)\chi^1(s)^\star}{\|s\|^{1+p}} ds \right\|_F,$$

where $\chi^1(s)$ *defined as in Lemma 1 with* $j = 1$.

Second, we study the asymptotic behavior of $\widehat{AT}_{wn}^F$ under the following alternative:

$$H_1^K : \text{MDDM}(e_t|Y_{t-1}) = \cdots = \text{MDDM}(e_t|Y_{t-K+1}) = 0, \text{MDDM}(e_t|Y_{t-K}) \neq 0,$$

for some $K \geq 1$.

Theorem 2. *Suppose that Assumptions 1(i)–(ii), 2–3 and 5 hold and as* $n \to \infty$. *Then, the test based on* $\widehat{AT}_{wn}^F$ *is consistent against* H_1^K, *for* $K \leq d$.

From Lemma 1, it is easy to obtain $\widehat{T}^F_{wn}(M) \to_d \sum_{j=1}^{M} \left\| \frac{1}{c_p} \int_{\mathcal{R}^p} \frac{\chi^j(s)\chi^j(s)^\star}{\|s\|^{1+p}} ds \right\|_F$ under the null hypothesis, and Assumptions 1–5 hold. Therefore, Theorem 1 is obtained because the residuals are MDS under the null hypothesis, and thus, $\text{MDDM}(e_t|Y_{t-j}) = 0$ for all $j \neq 0$. Hence, the optimal value for M is one, the minimum possible. Theorem 2 examines the power properties of our automatic test against the fixed alternative H_1^K. The proofs of both theorems are in the Supplementary Materials.

Remark 2. For simplicity, we consider the upper bound d to be a fixed large number. Similar to the automatic Portmanteau test (Escanciano and Labato, 2009), we can consider $d \equiv d(n)$ as an upper bound that grows slowly to infinity as $n \to \infty$. As pointed out by Escanciano and Labato (2009), considering $d \equiv d(n)$ can get a consistent test against all alternatives $H_1 : \text{MDDM}(e_t|Y_{t-j}) \neq 0$ for some $j \geq 1$ and Theorem 2 is true for all $K \geq 1$. And the disadvantages is that it may erroneously convey the interpretation of d as a bandwidth number because it grows slowly to infinity with the sample size. But the key difference for our number d and the bandwidth lies in that our number d has absolutely no influence in inference; for instance, it does not affect the convergence rate, contrary to Hong (1996). In Subsection 5.4, we briefly examine the sensitivity of the test to choose a value for d by simulations.

4. Bootstrap approximations

Given that the asymptotic null distribution of $\widehat{AT}^F_{wn}$ in Theorem 1 is not pivotal, its critical values have to be approximated. Similar to Wang et al. (2022), we apply a fixed-design wild bootstrap (WB) method to approximate the critical values of $\widehat{AT}^F_{wn}$. Our

fixed-design WB procedure is given as follows:

1. Estimate the original model and obtain the residuals $\{\widehat{e}_t\}_{t=1}^n$ according to $\widehat{e}_t = Y_t - f(\widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)$.

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 manner as that 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 $\widehat{AT}_{wn}^{F*} = \widehat{T}_{wn}^{F*}(M^*)$ in the same way as for $\widehat{T}_{wn}^{F}(M^*)$ with $\widehat{e}_t$ replaced by $\widehat{e}_t^{**}$, where M^* is defined as in Section 3.

6. Repeat steps 2–5 B times to obtain $\{\widehat{AT}_{wn,b}^{F*}\}_{b=1}^B$, and denote the αth upper percentile of $\{\widehat{AT}_{wn,b}^{F*}\}_{b=1}^B$ as c_α^*, which is considered as the approximated value of c_α in $\widehat{AT}_{wn}^{F} > c_\alpha$.

The distributions of $\{w_t^*\}$ are flexible in practice. And there are two popular selections for the distributions of $\{w_t^*\}$ such as:

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

and the Rademacher distribution

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

To justify the validity of our fixed-design WB method, we need one additional assumption, which is the same as the Assumption 6 in Wang et al.(2022) and is similar to Assumption A7 in Escanciano (2006). 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 $L^*(\widehat{\theta}_n) = E^*[\Upsilon(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)\Upsilon(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)^\top]$.

Assumption 6. *The estimator $\widehat{\theta}_n^*$ satisfies the asymptotic expansion*

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

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

(i) $E^*[\Upsilon(Y_t^*, \widehat{\mathcal{I}}_{t-1}, \widehat{\theta}_n)] = 0$ 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 $L(\theta_*) = E[\Upsilon(Y_t, \mathcal{I}_{t-1}, \theta_*)\Upsilon(Y_t, \mathcal{I}_{t-1}, \theta_*)^\top]$;

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

Lemma 2. *Suppose that Assumptions 1–3, 5 and 6 hold. Then, conditional on $\{Y_t, \widehat{\mathcal{I}}_{t-1}\}_{t=1}^n$,*

$$\widehat{AT}_{wn}^{F*} \to_d \left\| \frac{1}{c_p} \int_{\mathcal{R}^p} \frac{\chi_*^1(s)\chi_*^1(s)^\star}{\|s\|^{1+p}} ds \right\|_F \quad \text{in probability,}$$

where $\chi_*^j(s) = \Delta_*^j(s) - \phi_*^j(s)\mathcal{V}_*$, $\Delta_*^j(s)$, $\phi_*^j(s)$ and $\mathcal{V}_*$ are defined in the same way as $\Delta^j(s)$, $\phi^j(s)$ and $\mathcal{V}$ in Lemma 1 with θ_0 by θ_*.

Lemma 2 guarantees that our bootstrap critical value c_α^* computed from steps 1–6 is valid under the null hypothesis H_0 and any fixed alternative hypothesis H_1^K. In particular,

the above limit under H_0 is the same as the limiting null distribution of $\widehat{AT}^F_{wn}$, implying the asymptotic size accuracy. In addition,

$$\lim_{n\to\infty} P\left(\widehat{AT}^F_{wn} > c^*_\alpha\right) = 1 \text{ under } H_1^K,$$

meaning that $\widehat{AT}^F_{wn}$ with the bootstrapped critical value c^*_α can detect H_1^K consistently.

5. Simulation

In this section, we use simulation studies to investigate the performance of the proposed methods. We consider the dimensionality of the time series $p = 2$ and 5. First, we decide the tuning parameter k in the penalty term (3.1) in Subsection 5.1. Second, we compare our proposed methods with the other popular methods in finite sample. The details are presented in Subsections 5.2 and 5.3 for $p = 2$ and $p = 5$, respectively. Last, we show that all of the simulation results are not sensitive to the selected upper bound d in Subsection 5.4. For all simulations, we set the significance level $\alpha = 5\%$. For our proposed methods, the MDDM-based tests methods proposed by Wang et al. (2022) need the WB method to compute the critical value, we use the Rademacher distribution for w^*_t, and the bootstrap procedure is repeated $B = 1000$ times. The simulation results are based on 1000 replications.

5.1 Selection of k

The fixed number of k in (3.1) is a tuning parameter, which much be preselected. k determines which criterion, AIC or BIC, is selected. In this subsection, we show that

$k = 1.8$ is a suitable selection. Our null model is a vector AR(1) (VAR(1)) model:

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

To select k, we generate 1000 replications of sample size $n = 200, 1000$ from the following DGPs based on the above model (5.1):

$$\text{VAR}(1) : Y_t = 0.3Y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, I_p),$$

$$\text{VAR}(2) : Y_t = 0.3Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t,\ \varepsilon_t \sim N(0, I_p),$$

where VAR(1) is the null model and VAR(2) is the alternative model. For these experiments, we consider $p = 2, 5$ and $d = 15$. 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 adopt the least square estimate to obtain the parameter estimate $\widehat{A}_{0n}$ and $\widehat{A}_{1n}$. On the basis of the estimation, we calculate our data-driven MDDM-based test statistics $\widehat{AT}_{wn}^{F}$ and $\widehat{AT}_{wn}^{S}$ for different values of k.

Fig.1 shows the empirical rejection percentage (RP) of our proposed test at the 5% level for $k = 0, 0.1, \cdots, 3.3$, and the first row is corresponding to the size and the second row is corresponding to the power. From the size study, we know that the slop of the plot of the empirical RP becomes roughly flat for the values of k above 1.8. Moreover, it indicates that using a value of k larger than 1.8 is not necessary to control the type-I error properly. From the power study, we know that the power is flat for the values of k between 0 and 1.5 and decreases as k increases. We also consider all of the null and alternative models in Subsection 5.2, and all of the models show that the selection of $k = 1.8$ is a suitable choice. Hereafter, we fix $k = 1.8$ for all of the numerical studies.

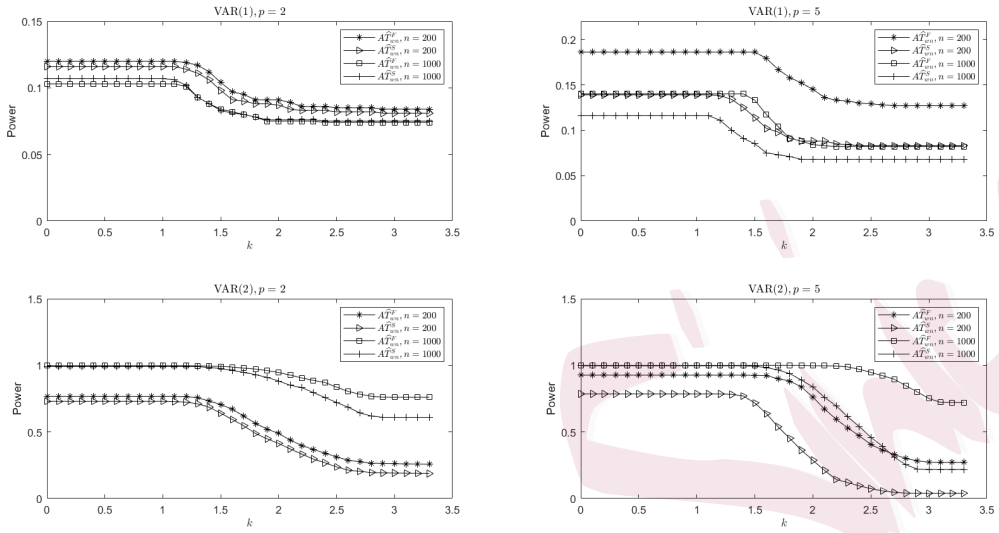

Figure 1: Rejection percentages (5% nominal level) of the tests $\widehat{AT}_{wn}^{F}$ and $\widehat{AT}_{wn}^{S}$ for the models VAR(1) and VAR(2) for several selected values of k and for $p = 2, 5$, where the first row is corresponding to the null model VAR(1) and the second row is corresponding to the alternative model VAR(2), respectively.

5.2 Finite sample tests comparison

In this subsection, we show the finite performance of the new data-driven tests $\widehat{AT}_{wn}^{F}$ and $\widehat{AT}_{wn}^{S}$ and compare it with the MDDM-based tests $\widehat{T}_{sn}^{F}(M), \widehat{T}_{sn}^{S}(M), \widehat{T}_{wn}^{F}(M)$ and $\widehat{T}_{wn}^{S}(M)$ for $M = 3, 6, 9$. In addition, we present some direct comparison with the tests in Box and Pierce (1970), Hosking (1980) and Lütkepohl (2005). The DGPs are the same as in Wang et al. (2022), which is also the multivariate counterparts of those univariate DGPs in Hong and Lee (2005) and Escanciano (2006).

5.2.1 Simulations for $p = 2$

In this subsection, we consider the simulation studies for $p = 2$. And the null model is also 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

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

On the basis of the model (5.2), we use the following two DGPs to examine the size performance of our tests:

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

$$\text{DGP 2} : \phi_1 = \phi_2 = \phi_5 = \phi_6 = 0.1, \phi_3 = \phi_4 = \phi_7 = 0.5,$$

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. ε_t is i.i.d. under DGP 1, while it has a multivariate ARCH structure under DGP 2.

To examine the power performance of our tests, we consider the following six DGPs:

$$\text{DGP 3} : Y_t = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{pmatrix} Y_{t-2} + \varepsilon_t;$$

$$\text{DGP 4} : Y_t = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{pmatrix} \varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP 5}: Y_t = \text{sign}(Y_{t-1}) + 0.43\varepsilon_t, \text{where sign}(x) = I(x > 0) - I(x < 0);$$

$$\text{DGP 6}: Y_{t,j} = \begin{cases} \varrho^{-1} Y_{t-1,j}, & \text{if } 0 \le Y_{t-1,j} < \varrho, \\ (1-\varrho)^{-1}(1 - Y_{t-1,j}), & \text{if } \varrho \le Y_{t-1,j} \le 1, \end{cases}$$

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

$$\text{DGP 7}: Y_t = \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \sin(0.3\pi Y_{t-2}) + \varepsilon_t;$$

$$\text{DGP 8}: Y_t = \begin{cases} \begin{pmatrix} 0.6 & -0.4 \\ 0.8 & 0.2 \end{pmatrix} Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} < 0, \\ \begin{pmatrix} -0.6 & 0.4 \\ -0.8 & -0.2 \end{pmatrix} Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} \ge 0, \end{cases}$$

where $\varepsilon_t = \eta_t$. The functions $\text{sign}(\cdot)$ and $\sin(\cdot)$ in DGPs 5 and 7 are evaluated elementwisely. With respect to the null VAR(1) model, we design the DGPs 3–4 as the linear models and DGPs 5–8 as the nonlinear models under the alternative.

For these experiments, we consider $n = 200, 1000, d = 15$ and $k = 1.8$. As in Subsection 5.1, we also use the least square to estimate the parameter $\widehat{\theta}_n = (vec(\widehat{A}_{0n})^\top, vec(\widehat{A}_{1n})^\top)^\top$, and obtain the model residual $\widehat{e}_t = Y_t - \widehat{A}_{0n} - \widehat{A}_{1n} Y_{t-1}$. On the basis of the $\{\widehat{e}_t\}$, we calculate our data-driven MDDM-based test statistics $\widehat{AT}_{wn}^F$ and $\widehat{AT}_{wn}^S$.

To make a comparison with correlation-based tests, we also compute the MDDM-based test statistics $\widehat{T}_{sn}^F(M)$, $\widehat{T}_{wn}^F(M)$, $\widehat{T}_{sn}^S(M), \widehat{T}_{wn}^S(M)(M = 3, 6, 9)$, 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

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

with $\widehat{C}_i = n^{-1}\sum_{t=i+1}^{n}\widehat{e}_t\widehat{e}_{t-i}^\top$, and the definition of $\widehat{LM}(M)$ can be found on p.172 of Lütkepohl (2005). At 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.

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

(1) Our proposed data-driven tests $\widehat{AT}^F_{wn}$ and $\widehat{AT}^S_{wn}$ and the MDDM-based tests $\widehat{T}^F_{wn}(M)$, $\widehat{T}^S_{wn}(M)$, $\widehat{T}^F_{sn}(M)$, and $\widehat{T}^S_{sn}(M)$ have a satisfactory size performance, although they are a little oversized when the sample size is $n = 200$; however, this case is relieved as the sample size increases, i.e., $n = 1000$.

(2) The Portmanteau tests $\widehat{Q}_1(M)$, $\widehat{Q}_2(M)$ and $\widehat{Q}_3(M)$ and the LM test $\widehat{LM}(M)$ perform as well as our proposed data-driven tests and the MDDM-based tests when the model error is homoskedastic as in DGP 1. While for the heteroskedasticity model error of DGP 2, the Portmanteau tests and LM test suffer from size distortion. The reason for this large size distortion is that the null distributions of all Portmanteau tests and the LM test are derived under the assumption that e_t is i.i.d., which does not hold for the heteroskedasticity model error of DGP 2.

From Table 1, we have the following findings for the power studies:

Table 1: The size and power (×100) of all tests for DGPs 1–8 at level 5%.

	DGP 1		DGP 2		DGP 3		DGP 4	
Test \ n	200	1000	200	1000	200	1000	200	1000
$\widehat{AT}^F_{wn}$	7.0	6.0	6.7	6.3	100	100	96.7	100
$\widehat{AT}^S_{wn}$	7.9	7.2	7.9	6.5	100	100	96.7	100
$\widehat{T}^F_{sn}(3)$	7.9	6.0	7.8	8.2	100	100	89.8	100
$\widehat{T}^F_{sn}(6)$	7.9	6.2	8.0	8.3	98.3	100	62.1	99.2
$\widehat{T}^F_{sn}(9)$	7.7	6.2	7.7	7.3	85.9	100	56.4	97.1
$\widehat{T}^S_{sn}(3)$	8.1	6.3	7.4	8.2	100	100	89.9	100
$\widehat{T}^S_{sn}(6)$	8.0	6.1	8.1	8.1	97.6	100	61.6	99.2
$\widehat{T}^S_{sn}(9)$	7.7	6.8	7.4	7.5	82.6	100	54.4	97.2
$\widehat{T}^F_{wn}(3)$	7.2	6.0	7.5	8.5	100	100	87.4	100
$\widehat{T}^F_{wn}(6)$	7.1	6.9	7.9	7.9	97.4	100	60.9	99.1
$\widehat{T}^F_{wn}(9)$	6.3	6.8	7.2	7.4	83.7	100	54.8	96.4
$\widehat{T}^S_{wn}(3)$	7.4	5.9	7.3	8.6	100	100	86.9	99.9
$\widehat{T}^S_{wn}(6)$	6.6	7.2	7.7	7.9	97.1	100	60.6	99.1
$\widehat{T}^S_{wn}(9)$	6.6	6.3	7.3	7.5	82.7	100	54.1	96.2
$\widehat{Q}_1(3)$	5.8	5.1	11.1	11.3	100	100	100	100
$\widehat{Q}_1(6)$	5.3	4.5	10.8	10.1	100	100	99.9	100
$\widehat{Q}_1(9)$	4.3	4.5	10.7	9.0	100	100	99.7	100
$\widehat{Q}_2(3)$	6.5	5.2	11.4	11.7	100	100	100	100
$\widehat{Q}_2(6)$	6.0	5.0	11.6	10.5	100	100	100	100
$\widehat{Q}_2(9)$	6.3	4.9	11.7	9.4	100	100	99.9	100
$\widehat{Q}_3(3)$	6.2	5.1	11.3	11.4	100	100	100	100
$\widehat{Q}_3(6)$	5.7	4.7	11.5	10.4	100	100	100	100
$\widehat{Q}_3(9)$	6.6	4.8	11.9	9.3	100	100	99.9	100
$\widehat{LM}(3)$	4.9	4.3	11.1	9.9	100	100	100	100
$\widehat{LM}(6)$	4.2	4.5	9.1	10.4	100	100	100	100
$\widehat{LM}(9)$	3.2	3.9	6.3	7.8	100	100	100	100
	DGP 5		DGP 6		DGP 7		DGP 8	
Test \ n	200	1000	200	1000	200	1000	200	1000
$\widehat{AT}^F_{wn}$	98.2	100	100	100	99.6	100	100	100
$\widehat{AT}^S_{wn}$	98.2	100	100	100	99.5	100	100	100
$\widehat{T}^F_{sn}(3)$	98.0	100	100	100	94.1	100	62.7	99.2
$\widehat{T}^F_{sn}(6)$	97.7	100	55.6	99.9	49.8	96.8	26.9	68.7
$\widehat{T}^F_{sn}(9)$	97.4	100	25.8	71.0	30.9	81.3	20.2	46.6
$\widehat{T}^S_{sn}(3)$	98.0	100	100	100	93.6	100	57.3	99.0
$\widehat{T}^S_{sn}(6)$	97.6	100	40.4	96.4	47.4	97.0	24.1	65.0
$\widehat{T}^S_{sn}(9)$	97.1	100	20.2	53.6	28.1	80.3	19.0	43.3
$\widehat{T}^F_{wn}(3)$	97.9	100	100	100	97.0	100	93.2	100
$\widehat{T}^F_{wn}(6)$	97.6	100	100	100	70.6	97.7	71.8	100
$\widehat{T}^F_{wn}(9)$	97.5	100	100	100	49.2	98.4	58.4	99.0
$\widehat{T}^S_{wn}(3)$	97.9	100	100	100	97.0	100	90.3	100
$\widehat{T}^S_{wn}(6)$	97.6	100	100	100	69.3	97.7	68.4	99.8
$\widehat{T}^S_{wn}(9)$	97.4	100	100	100	48.8	98.2	56.0	98.8
$\widehat{Q}_1(3)$	95.8	100	6.6	4.8	71.3	98.7	13.0	28.0
$\widehat{Q}_1(6)$	94.8	100	6.3	5.6	58.8	96.9	10.0	18.9
$\widehat{Q}_1(9)$	93.2	100	6.2	5.5	50.7	94.0	9.6	15.4
$\widehat{Q}_2(3)$	95.9	100	6.9	4.9	71.7	98.7	13.9	28.4
$\widehat{Q}_2(6)$	95.0	100	7.0	5.9	60.1	97.1	10.6	19.7
$\widehat{Q}_2(9)$	94.4	100	8.1	6.1	55.3	94.1	12.6	16.6
$\widehat{Q}_3(3)$	95.9	100	6.9	4.8	71.7	98.7	13.5	28.3
$\widehat{Q}_3(6)$	95.0	100	7.0	5.7	60.0	97.1	10.5	19.7
$\widehat{Q}_3(9)$	94.3	100	7.2	6.0	54.5	94.1	12.0	16.3
$\widehat{LM}(3)$	97.3	100	2.2	1.9	66.1	98.5	7.1	19.3
$\widehat{LM}(6)$	96.4	100	2.6	3.2	49.8	95.4	6.8	14.1
$\widehat{LM}(9)$	94.8	100	2.7	2.7	36.5	91.3	5.7	11.4

(1) For all of the DGPs 3–8, our proposed data-driven tests $\widehat{AT}_{wn}^{F}$ and $\widehat{AT}_{wn}^{S}$ have a considerable power. For most of the examined case as in DGPs 3–8, the power for the tests $\widehat{AT}_{wn}^{F}$ and $\widehat{AT}_{wn}^{S}$ tends to 1.0, and the smallest power is more than 0.96 for DGP 4 with $n = 200$.

(2) In most examined cases, the power of all $\widehat{T}_{sn}^{F}(M), \widehat{T}_{sn}^{S}(M), \widehat{T}_{wn}^{F}(M), \widehat{T}_{wn}^{S}(M), \widehat{Q}_i(M)$, and $\widehat{LM}(M)$ decreases as the value of M increases. For the linear alternative models (i.e., DGPs 3 and 4), the data-driven method has similar performance as the correlation-based tests $\widehat{Q}_i(M)$ and $\widehat{LM}(M)$, while the MDDM-based tests have inferior power performance when $n = 200$, but their power at $n = 1000$ performs as well as our proposed data-driven tests. For the nonlinear alternative models (i.e., DGPs 5-8), our proposed tests and all MDDM-based tests in general are much more powerful than the correlation-based tests $\widehat{Q}_i(M)$ and $\widehat{LM}(M)$, especially when the sample size $n = 200$. In particular, the power of $\widehat{Q}_i(M)$ and $\widehat{LM}(M)$ is very low for DGPs 6 and 8.

5.2.2 Simulations for $p = 5$

In this subsection, we consider the simulation studies when the dimensionality of Y_t is $p = 5$. Similar to the previous, 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

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

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):

$$\text{DGP 9 :}\phi_1 = 1 \text{ and } \phi_2 = \phi_3 = \phi_4 = 0;$$

$$\text{DGP 10 :}\phi_1 = \phi_3 = 0.1 \text{ and } \phi_2 = \phi_4 = 0.5,$$

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:

$$\text{DGP 11 :}Y_t = 0.3Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t;$$

$$\text{DGP 12 :}Y_t = 0.3Y_{t-1} + 0.3\varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP 13 :}Y_t = \text{sign}(Y_{t-1}) + 0.43\varepsilon_t;$$

$$\text{DGP 14 :}Y_{t,j} = \begin{cases} \varrho^{-1}Y_{t-1,j}, & \text{if } 0 \le Y_{t-1,j} < \varrho, \\ (1-\varrho)^{-1}(1-Y_{t-1,j}), & \text{if } \varrho \le Y_{t-1,j} \le 1, \end{cases}$$

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

$$\text{DGP 15 :}Y_t = 0.3Y_{t-1} + 0.3\sin(0.3\pi Y_{t-2}) + \varepsilon_t;$$

$$\text{DGP 16 :}Y_t = \begin{cases} 0.3Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} < 0, \\ -0.3Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1,1} \ge 0, \end{cases}$$

where $\varepsilon_t = \eta_t$. For each replication, we compute all MDDM-based tests and correlation-based tests as performed in Subsection 5.2.1.

Table 2 reports the size and power of all examined tests. The findings from this table are qualitatively similar to those from Table 1. In terms of size, there is quite a

Table 2: The size and power (×100) of all tests for DGPs 9–16 at level 5%.

	DGP 9		DGP 10		DGP 11		DGP 12	
Test \ n	200	1000	200	1000	200	1000	200	1000
$\widehat{AT}^F_{wn}$	10.1	8.1	6.7	6.6	90.4	100	75.5	97.2
$\widehat{AT}^S_{wn}$	8.3	6.9	6.8	6.5	54.8	100	31.6	53.5
$\widehat{T}^F_{sn}(3)$	11.1	7.7	7.8	6.7	89.6	100	68.0	97.3
$\widehat{T}^F_{sn}(6)$	12.6	8.4	7.9	6.9	72.6	99.2	51.2	80.6
$\widehat{T}^F_{sn}(9)$	16.3	8.8	7.8	7.4	67.0	97.6	48.7	71.0
$\widehat{T}^S_{sn}(3)$	8.9	6.5	7.9	6.6	71.2	99.0	50.5	80.8
$\widehat{T}^S_{sn}(6)$	8.6	7.0	7.9	6.6	50.2	89.4	33.8	51.4
$\widehat{T}^S_{sn}(9)$	8.2	6.6	7.4	6.7	43.7	78.9	29.6	40.9
$\widehat{T}^F_{wn}(3)$	11.2	7.6	8.0	7.0	86.3	100	63.7	95.9
$\widehat{T}^F_{wn}(6)$	12.8	8.5	8.6	7.2	67.4	100	50.2	76.5
$\widehat{T}^F_{wn}(9)$	14.9	9.4	8.5	6.1	62.4	100	47.6	64.6
$\widehat{T}^S_{wn}(3)$	9.0	6.5	7.8	7.0	71.3	99.6	50.9	83.2
$\widehat{T}^S_{wn}(6)$	10.4	7.8	8.2	6.7	54.8	99.5	39.4	59.3
$\widehat{T}^S_{wn}(9)$	12.0	7.9	8.1	6.3	50.5	99.1	38.7	48.0
$\widehat{Q}_1(3)$	5.7	4.9	6.3	6.3	86.2	100	67.0	99.9
$\widehat{Q}_1(6)$	5.1	5.0	9.1	7.2	71.2	99.8	51.5	96.3
$\widehat{Q}_1(9)$	7.0	5.5	10.7	6.5	66.9	99.3	51.4	91.5
$\widehat{Q}_2(3)$	7.5	5.3	7.5	7.5	87.5	100	68.7	99.9
$\widehat{Q}_2(6)$	7.6	5.6	12.6	7.5	75.6	99.8	59.2	96.8
$\widehat{Q}_2(9)$	12.6	8.1	18.2	7.4	76.2	99.7	61.5	92.8
$\widehat{Q}_3(3)$	7.4	5.3	7.5	9.7	87.5	100	68.9	99.9
$\widehat{Q}_3(6)$	7.1	5.5	12.3	7.5	75.4	99.8	58.4	96.8
$\widehat{Q}_3(9)$	12.1	7.6	16.6	9.4	75.6	99.7	60.8	92.8
$\widehat{LM}(3)$	4.0	5.7	4.6	6.8	67.3	100	60.0	99.2
$\widehat{LM}(6)$	3.4	3.9	4.0	4.1	42.5	99.5	31.7	92.2
$\widehat{LM}(9)$	3.7	4.1	3.3	3.4	30.6	97.5	21.4	77.6
	DGP 13		DGP 14		DGP 15		DGP 16	
Test \ n	200	1000	200	1000	200	1000	200	1000
$\widehat{AT}^F_{wn}$	100	100	100	100	73.7	99.7	84.0	99.6
$\widehat{AT}^S_{wn}$	100	100	100	100	34.1	75.8	58.5	95.1
$\widehat{T}^F_{sn}(3)$	100	100	71.5	100	70.5	99.9	13.3	13.3
$\widehat{T}^F_{sn}(6)$	100	100	30.8	58.7	55.6	92.1	15.6	11.1
$\widehat{T}^F_{sn}(9)$	100	100	25.2	31.5	49.8	85.4	17.0	12.3
$\widehat{T}^S_{sn}(3)$	100	100	38.1	96.5	53.2	94.8	10.4	9.8
$\widehat{T}^S_{sn}(6)$	100	100	16.4	29.0	36.2	72.6	11.5	9.4
$\widehat{T}^S_{sn}(9)$	100	100	13.8	16.3	32.6	62.6	12.1	8.4
$\widehat{T}^F_{wn}(3)$	100	100	100	100	69.2	99.8	17.3	23.8
$\widehat{T}^F_{wn}(6)$	100	100	95.2	100	53.9	91.8	17.9	18.8
$\widehat{T}^F_{wn}(9)$	100	100	88.8	100	48.9	83.4	18.3	17.0
$\widehat{T}^S_{wn}(3)$	100	100	95.0	100	55.6	94.7	14.7	16.2
$\widehat{T}^S_{wn}(6)$	100	100	79.0	100	41.1	78.3	15.1	14.3
$\widehat{T}^S_{wn}(9)$	100	100	69.7	100	38.5	67.8	15.3	12.6
$\widehat{Q}_1(3)$	99.3	100	6.1	7.0	63.0	99.8	7.3	6.1
$\widehat{Q}_1(6)$	98.7	100	6.4	5.6	45.8	97.2	5.8	6.3
$\widehat{Q}_1(9)$	98.3	100	7.9	6.0	47.2	92.9	8.7	8.1
$\widehat{Q}_2(3)$	99.3	100	6.8	7.6	64.7	99.8	7.7	6.5
$\widehat{Q}_2(6)$	98.9	100	9.0	6.6	52.6	97.8	8.0	7.1
$\widehat{Q}_2(9)$	99.0	100	13.6	7.1	57.4	93.7	14.1	9.1
$\widehat{Q}_3(3)$	99.4	100	6.8	7.4	64.7	99.8	7.6	6.5
$\widehat{Q}_3(6)$	98.9	100	8.9	6.3	52.0	97.8	7.9	6.7
$\widehat{Q}_3(9)$	99.0	100	12.6	7.0	56.7	93.8	13.3	8.9
$\widehat{LM}(3)$	99.7	100	3.1	1.9	40.1	99.2	4.2	3.3
$\widehat{LM}(6)$	99.7	100	2.7	2.9	22.1	90.2	4.4	4.2
$\widehat{LM}(9)$	99.1	100	3.2	4.0	17.1	76.8	4.3	5.8

bit of size distortion with the Frobenius norm-based tests when $n = 200$ for DGP 9, and the size distortion noticeably reduces at sample size $n = 1000$. In terms of power, our tests 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.

5.3 Simulations for a high-order dependent: VAR(10)

To demonstrate that our proposed data-driven tests have good performance in the high-order dependent, we consider the following DGP:

$$\text{VAR}(10) : Y_t = 0.3Y_{t-1} + \beta Y_{t-10} + \varepsilon_t,\ \varepsilon_t \sim N(0, I_p).$$

Our null model is VAR(1):

$$Y_t = A_0 + A_1 Y_{t-1} + \varepsilon_t. \tag{5.4}$$

For each replication, we also compute all MDDM-based tests and correlation-based tests as performed in Subsection 5.2.

Table 3 reports the empirical RP for six values of $\beta = -0.4, -0.3, -0.2, 0.2, 0.3$, and 0.4. Moreover, we consider sample size $n = 1000$ and $p = 2, 5$. From Table 3, we have the following findings for the study:

(1) The emprical RP for all tests increases as the absolute value of β increases. The reason for this is that the dependent increases as the absolute value of β increases. For all the considered six cases and two considered $p = 2, 5$ values, the power of the MDDM-based tests $\widehat{T}^F_{sn}(M), \widehat{T}^S_{sn}(M), \widehat{T}^F_{wn}(M), \widehat{T}^S_{wn}(M)$ and the power of the correlation-based tests

Table 3: Empirical power (percentages) for VAR(10) for different values of β and the sample size is $n = 1000$.

p	β	-0.4	-0.3	-0.2	0.2	0.3	0.4
2	$\widehat{AT}^F_{wn}$	100	85.5	13.3	13.7	83.5	100
	$\widehat{AT}^S_{wn}$	100	68.4	11.5	11.6	64.8	99.8
	$\widehat{T}^F_{sn}(3)$	22.3	14.0	9.8	7.7	16.3	26.3
	$\widehat{T}^F_{sn}(6)$	24.4	14.5	10.3	15.8	27.2	50.3
	$\widehat{T}^F_{sn}(9)$	84.4	47.1	21.9	26.2	56.1	87.8
	$\widehat{T}^S_{sn}(3)$	22.0	13.8	10.7	7.6	15.1	24.2
	$\widehat{T}^S_{sn}(6)$	20.0	12.8	9.5	14.1	24.7	45.2
	$\widehat{T}^S_{sn}(9)$	77.6	41.2	19.7	21.8	49.8	83.0
	$\widehat{T}^F_{wn}(3)$	20.9	13.4	9.3	7.2	16.0	24.8
	$\widehat{T}^F_{wn}(6)$	23.6	15.0	9.6	15.2	26.8	49.8
	$\widehat{T}^F_{wn}(9)$	80.9	44.8	20.1	23.7	53.8	84.7
	$\widehat{T}^S_{wn}(3)$	20.9	12.7	8.8	7.4	16.4	24.3
	$\widehat{T}^S_{wn}(6)$	23.2	14.7	9.5	14.9	26.7	48.5
	$\widehat{T}^S_{wn}(9)$	78.6	42.4	20.3	23.5	51.5	83.4
	$\widehat{Q}_1(3)$	19.9	11.9	8.0	6.5	10.5	17.6
	$\widehat{Q}_1(6)$	25.4	14.2	7.6	13.2	19.9	41.4
	$\widehat{Q}_1(9)$	29.0	16.2	8.6	12.7	20.2	41.1
	$\widehat{Q}_2(3)$	20.3	12.2	8.2	6.6	10.8	18.0
	$\widehat{Q}_2(6)$	26.2	14.7	8.2	13.4	21.0	42.2
	$\widehat{Q}_2(9)$	30.9	17.1	9.1	13.7	21.3	42.2
	$\widehat{Q}_3(3)$	20.2	12.1	8.2	6.6	10.8	18.0
	$\widehat{Q}_3(6)$	26.0	14.3	7.9	13.4	20.7	41.7
	$\widehat{Q}_3(9)$	30.5	17.1	9.0	13.0	21.2	42.0
	$\widehat{LM}(3)$	20.4	11.2	7.4	7.9	12.3	23.1
	$\widehat{LM}(6)$	21.8	9.9	5.7	8.5	14.8	34.7
	$\widehat{LM}(9)$	28.8	12.6	5.9	8.8	15.6	35.3
5	$\widehat{AT}^F_{wn}$	100	98.7	17.7	20.5	97.7	100
	$\widehat{AT}^S_{wn}$	100	92.6	16.2	17.5	92.4	100
	$\widehat{T}^F_{sn}(3)$	55.6	39.9	15.6	15.7	52.7	69.4
	$\widehat{T}^F_{sn}(6)$	73.6	58.6	20.3	31.6	82.0	93.8
	$\widehat{T}^F_{sn}(9)$	99.7	96.6	41.3	49.7	99.0	99.9
	$\widehat{T}^S_{sn}(3)$	43.9	32.5	13.1	13.3	39.0	51.1
	$\widehat{T}^S_{sn}(6)$	55.3	39.6	13.7	22.3	61.9	77.2
	$\widehat{T}^S_{sn}(9)$	93.7	83.9	26.1	31.0	90.6	96.8
	$\widehat{T}^F_{wn}(3)$	53.9	37.7	15.3	16.9	49.9	67.6
	$\widehat{T}^F_{wn}(6)$	74.0	59.2	20.4	31.1	80.8	92.1
	$\widehat{T}^F_{wn}(9)$	99.6	95.7	40.4	45.9	98.5	99.9
	$\widehat{T}^S_{wn}(3)$	46.2	33.6	14.2	14.4	41.3	57.2
	$\widehat{T}^S_{wn}(6)$	68.0	51.2	17.6	26.3	71.6	87.3
	$\widehat{T}^S_{wn}(9)$	98.5	91.2	34.4	38.4	95.8	99.2
	$\widehat{Q}_1(3)$	52.1	37.0	11.7	13.1	39.2	53.2
	$\widehat{Q}_1(6)$	77.4	62.4	18.9	24.6	75.6	90.8
	$\widehat{Q}_1(9)$	83.8	67.3	23.3	29.3	79.4	90.7
	$\widehat{Q}_2(3)$	53.4	38.3	12.2	14.0	40.2	54.1
	$\widehat{Q}_2(6)$	79.5	64.4	20.5	26.7	77.0	91.6
	$\widehat{Q}_2(9)$	85.3	70.6	26.2	32.5	82.3	92.3
	$\widehat{Q}_3(3)$	53.2	38.2	12.0	14.0	40.1	54.0
	$\widehat{Q}_3(6)$	79.0	64.0	20.2	26.6	76.9	91.5
	$\widehat{Q}_3(9)$	85.3	70.3	26.1	31.9	82.0	92.2
	$\widehat{LM}(3)$	70.5	52.7	15.3	15.2	55.2	74.6
	$\widehat{LM}(6)$	79.4	61.8	13.6	15.6	72.4	88.0
	$\widehat{LM}(9)$	100	100.0	72.7	66.4	100	100

$\widehat{Q}_i(M)$ and $\widehat{LM}(M)$ increase as M increases.

(2) For $p = 2$, our proposed data-driven tests perform better than all of the MDDM-based tests and the correlation-based tests for $\theta = -0.4, -0.3, 0.3$, and 0.4. While for $\theta = -0.2$ and 0.2, the data-driven tests perform similarly as $\widehat{T}^F_{sn}(6)$, $\widehat{T}^F_{sn}(9)$, $\widehat{T}^F_{wn}(6)$, $\widehat{T}^F_{wn}(9)$, $\widehat{Q}_i(6)$, and $\widehat{Q}_i(9)$ for $i = 1, 2, 3$ and has better performance than $\widehat{T}^F_{sn}(3)$, $\widehat{T}^F_{wn}(3)$ and the LM tests $\widehat{LM}(3)$, $\widehat{LM}(6)$, and $\widehat{LM}(9)$. Moreover, the MDDM-based methods $\widehat{T}^F_{sn}(9)$ and $\widehat{T}^F_{wn}(9)$ have similar performance and perform better than the other MDDM-based methods.

(3) For $p = 5$, we can obtain similar results as $p = 2$. Given that this case is suitable for LM tests, $\widehat{LM}(9)$ has the best performance among all the considered test methods. By contrast, our proposed methods $\widehat{AT}^F_{wn}$ and $\widehat{AT}^S_{wn}$ perform as well as $\widehat{LM}(9)$, $\widehat{T}^F_{sn}(9)$, $\widehat{T}^F_{wn}(9)$, $\widehat{T}^S_{sn}(9)$, and $\widehat{T}^S_{wn}(9)$ methods and perform better than the other methods.

5.4 Selection of d

In this subsection, we examine the sensitivity of the tests to the selection of the upper bound d. Similarly, the null model and DGPs used are consistent with those in Section 5.1.

Table 4: RP (percentages) of the data-deriven test for different values of d with nominal0.05 and $n = 1000$.

	AR(1),$p = 2$		AR(1),$p = 5$		AR(2),$p = 2$		AR(2),$p = 5$	
d	$A\widehat{T}^F_{wn}$	$A\widehat{T}^S_{wn}$	$A\widehat{T}^F_{wn}$	$A\widehat{T}^S_{wn}$	$A\widehat{T}^F_{wn}$	$A\widehat{T}^S_{wn}$	$A\widehat{T}^F_{wn}$	$A\widehat{T}^S_{wn}$
15	7.0	7.0	9.1	6.9	98.2	94.2	100	92.5
20	7.0	7.0	9.8	6.9	98.3	94.6	100	92.7
25	7.0	7.0	10.4	6.9	98.4	94.9	100	92.9
30	7.0	7.0	10.8	7.1	98.4	95.5	100	93.3
35	7.0	7.0	11.2	7.1	98.4	95.5	100	93.6
40	7.0	7.0	11.5	7.1	98.4	95.5	100	93.9

Table 4 reports the result for $n = 1000$ and six values of $d = 15, 20, 25, 30, 35$ and 40, which shows that the proposed test is completely insensitive to the choice of d. We perform additional experiments under the null and alternative conditions for various sample sizes and model specifications, and all cases show that the results is absolute lack of sensitivity to the selection of d.

6. Real data example

In the real data analysis, we apply our proposed data-driven MDDM-based methods to analyze the dataset of U.S. monthly interest rates. The time series spans from 1959.1 to 1993.2 and includes two components: three month treasury bills and three year treasury notes. These components represent short-term and intermediate series, respectively, in the term structure of interest rates. We denote the interest-rate series as $IR_t = (IR_{1t}, IR_{2t})^\top$. To analyze the growth patterns, we introduce the growth series $Y_t = (Y_{1t}, Y_{2t})^\top$, which comprises 409 observations. Specifically, $Y_{it} = \log(IR_{it}) - \log(IR_{i,t-1})$ for $i = 1, 2$ represents the growth rate of the ith interest rate component at time t. Here, Y_{it} denotes the interest rate component i at time t.

Tsay (1998) utilized a three-regime threshold vector autoregressive (TVAR) model to analyze the dataset $\{Y_t\}_{t=1}^{409}$, where

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

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

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

where $X_t = \log(IR_{1t}) - \log(IR_{2t})$ is the three month "average spread" in logged interest rates.

In model (6.1), the unknown parameters $A_i^{(j)}$ are estimated using least squares estimates, and the error term ε_t is assumed to have a threshold constant variance structure. In accordance with Assumption 1b in Tsay (1998), a multivariate martingale difference assumption is made, stating that $E(\varepsilon_t|\mathcal{I}_{t-1}) = 0$, but it is not tested for model (6.1). To examine the multivariate MDH for ε_t in model (6.1), we applied our proposed data-driven tests and MDDM-based tests. The results are presented in Table 5. The table provides strong evidence from all the MDDM-based tests and our proposed tests, indicating that ε_t in model (6.1) satisfies the multivariate MDH.

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

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

or a VAR(m) model:

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

where the order m 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 our data-driven tests and the MDDM-based tests to examine the MDH for ε_t in models (6.2)-(6.3), and the results are given in Table 5.

Table 5: The p-values of data-driven tests and MDDM-based tests at lag $M = 1, ..., 10$ for five different models.

Tests		TVAR	Constant	VAR(2)	VAR(6)	VAR(7)
$\widehat{AT}^F_{wn}$		**0.650**	0.000	0.005	0.011	0.005
$\widehat{AT}^S_{wn}$		**0.645**	0.000	0.006	0.011	0.005
$\widehat{T}^F_{sn}(M)$	1	**0.656**	0.000	0.046	**0.060**	**0.061**
	2	**0.847**	0.000	0.019	0.016	0.010
	3	**0.787**	0.000	0.025	0.004	0.000
	4	**0.845**	0.000	**0.073**	0.009	0.003
	5	**0.865**	0.000	**0.122**	0.013	0.007
	6	**0.896**	0.000	0.022	0.013	0.009
	7	**0.889**	0.000	0.023	0.014	0.011
	8	**0.879**	0.000	0.028	0.020	0.025
	9	**0.879**	0.000	0.045	0.040	0.035
	10	**0.901**	0.000	**0.068**	**0.066**	0.025
$\widehat{T}^S_{sn}(M)$	1	**0.652**	0.000	0.046	**0.060**	**0.061**
	2	**0.840**	0.000	0.019	0.016	0.010
	3	**0.779**	0.000	0.025	0.003	0.000
	4	**0.837**	0.000	**0.073**	0.009	0.003
	5	**0.865**	0.000	**0.123**	0.013	0.007
	6	**0.893**	0.000	0.022	0.013	0.009
	7	**0.888**	0.000	0.024	0.015	0.011
	8	**0.877**	0.000	0.029	0.023	0.025
	9	**0.875**	0.000	0.045	0.040	0.035
	10	**0.899**	0.000	**0.069**	**0.066**	0.025
$\widehat{T}^F_{wn}(M)$	1	**0.650**	0.000	0.046	**0.060**	**0.061**
	2	**0.735**	0.000	0.021	0.011	0.005
	3	**0.732**	0.000	0.034	0.010	0.006
	4	**0.745**	0.000	**0.083**	0.011	0.009
	5	**0.751**	0.000	0.029	0.002	0.004
	6	**0.765**	0.000	0.005	0.005	0.005
	7	**0.766**	0.000	0.004	0.007	0.008
	8	**0.763**	0.000	0.006	0.014	0.005
	9	**0.763**	0.000	0.009	0.014	0.009
	10	**0.770**	0.000	0.014	0.024	0.017
$\widehat{T}^S_{wn}(M)$	1	**0.645**	0.000	0.046	**0.060**	**0.061**
	2	**0.729**	0.000	0.021	0.011	0.005
	3	**0.727**	0.000	0.034	0.010	0.006
	4	**0.742**	0.000	**0.084**	0.011	0.009
	5	**0.750**	0.000	0.030	0.002	0.004
	6	**0.763**	0.000	0.006	0.005	0.005
	7	**0.763**	0.000	0.004	0.007	0.008
	8	**0.762**	0.000	0.009	0.014	0.005
	9	**0.761**	0.000	0.009	0.014	0.009
	10	**0.767**	0.000	0.013	0.024	0.017

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

According to this table, all MDDM-based tests and our proposed data-driven tests strongly reject the constant model, indicating that the interest rate market is not efficient. Additionally, most MDDM-based tests and our tests reject the MDH in all three VAR(m) models at a 5% significance level, suggesting that the VAR(m) model in (6.3) does not adequately fit the data $\{Y_t\}_{t=1}^{409}$. We did not consider the LM test $\widehat{LM}(M)$ and the Portmanteau tests $\widehat{Q}_i(M)$ $(i = 1, 2, 3)$ for models (6.1)–(6.3) because their errors $\{\varepsilon_t\}$ are not independent and identically distributed. Furthermore, we observe conflicting results among fixed-order MDDM tests for different lag lengths when analyzing data from VAR(m) models with $m = 2, 6, 7$.

Overall, our test results support the use of the TVAR model (6.1) to fit this bivariate exchange rate dataset. The dataset exhibits a clear threshold effect, which cannot be adequately captured by the constant and linear VAR models. Therefore, the TVAR model is a more suitable choice in this case.

7. Concluding and Discussion

In this article, we propose a data-driven MDDM-based test for the multivariate MDH in stationary time series models. Unlike the MDDM-based tests proposed by Wang et al. (2022) need to specify the lag order in prior, the data-driven MDDM-based test can automatically select the lag order by the data. The data selects whether the BIC or the AIC criterion is employed in the selection of the lag order M. Compared with the existing test methods, the data-driven tests have additional interesting advantages. First, under the null hypothesis, the lag order is one, which is simple to implement. Second,

the data-driven test exhibits higher empirical power in finite samples than the rival ones, especially in detecting the model inadequacy caused by high-order dependence. Last, the data-driven test is particularly suitable for financial data because it can detect both of linear and non-linear dependence.

There are some possible extension for the current study. One possible direction is to consider the robust MDDM-based test or the robust data-driven test. Another possible direction, the procedures developed here could be extended to explore the data-driven selection of lag M for alternative test statistics, such as the one proposed by Mehta et al. (2019). This move would provide a broader framework for selecting the appropriate order in various testing scenarios.

Supplementary Materials

The supplementary materials include some additional simulation results as well as the proofs of all theorems and lemmas in the paper.

Acknowledgments

The authors greatly appreciate the very helpful comments and suggestions of two anonymous referees, associate editor, and co-editor. This work is partially supported by the National Science Foundation of China (No.12271213), Guangdong Basic and Applied Basic Research Foundation (No.32224156).

References

Akaike H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control **94**, 869-879.

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.

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.

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

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

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.

Escanciano, J, C. and Lobato, I, N. (2009). An automatic portmanteau test for serial correlation. Journal of Econometrics **151**, 140-149.

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

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

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.

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

Inglot, T. and Ledwina, T. (2006a). Towards data driven selection of a penalty function for data driven Neyman tests. Linear algebra and its applications **1**, 124-133.

Inglot, T. and Ledwina, T. (2006b). Data-driven score tests for homoscedastic linear regression model: asymptotic results. Probability and mathematical statistics-wroclaw university **26**, 41.

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.

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.

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

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

Lobato, I., Nankervis, J.C. and Savin, N. E. (2001). Testing for autocorrelation using a modified box-pierce Q test. International Economic Review, **42**, 187-205.

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

Mehta, R., Chung, J., Shen, C., Xu, T. and Vogelstein, J. T. (2019). Independence Testing for Multivariate Time Series. arXiv:1908.06486.

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

Schwarz G. (1978). Estimating the dimension of a model. The annals of statistics **6**, 461-464.

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

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.

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.

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

Wang, G., Zhu, K. and Shao, X. (2022). Testing for the martingale difference hypothesis in multivariate time series models. Journal of Business & Economic Statistics **40**, 980-994.

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

Woodroofe, M. (1982). On model selection and the arc sine laws. The Annals of Statistics **10**, 1182-1194.

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

School of Economics, Jinan University, Guangzhou, 510632, China

E-mail: planckzero@163.com

School of Economics, Jinan University, Guangzhou, 510632, China

E-mail: twanggc@jnu.edu.cn.