Statistica Sinica **36** (2026) 1-24
doi:https://doi.org/10.5705/ss.202024.0035

SUPPLEMENTARY MATERIAL FOR "AN AUTOMATIC MDDM-BASED TEST FOR MARTINGALE DIFFERENCE HYPOTHESIS"

The supplementary materials include the finite sample performance for the case when the dimension of the time series is $p = 1$, the finite sample performance when the moment conditions fail in Assumption 1, and the proofs of all theorems and lemmas in the paper. The finite sample performance for $p = 1$ is presented in Section A, the finite sample performance when the moment conditions fail in Assumption 1 is given in Section B, and the Section C provides the proofs of the theorems and lemmas.

A. Simulations for $p = 1$

In this section, we use the simulation studies to investigate the performance of the proposed methods for $p = 1$. These studies are divided into three parts: the first part is to demonstrate that the tuning parameter $k = 1.8$ is also suitable for $p = 1$ (See Section A.1). The second part is to investigate the performance of the proposed methods for varied DGPs (See Section A.2) and to investigate the performance for the high dependent is presented in Section A.3. The third part aims to verify that the proposed testing methods are not sensitivity to the selected values of d (See Section A.4). For all simulations, we set the significance level $\alpha = 5\%$. For our proposed methods, the MDDM-based tests methods and the methods proposed by Escanciano (2006) need the wild bootstrap method to compute the critical value and we use the Rademacher distribution for w_t^* and the bootstrap procedure are repeated $B = 1000$ times. The simulation results are based on 1000 replications. For $p = 1$, the spectrum norm and Frobenius norm is equivalent, then, we only need to consider the Frobenius norm, i.e., we only consider the testing statistic $\widehat{AT}_{wn}^{F}$ for our proposed data-driven method and the MDDM-based methods $\widehat{T}_{wn}^F(M)$ and $\widehat{T}_{sn}^F(M)$ proposed by Wang, Zhu and Shao (2022).

A.1 Selection of k

In this subsection, we use simulation studies to verify that the tuning parameter $k = 1.8$ is also suitable for $p = 1$. The null model is the AR(1) model:

$$Y_t = a_0 + a_1 Y_{t-1} + \varepsilon_t. \tag{A.1}$$

To select k, we generate 1000 replications of sample size $n = 200, 1000$ from the following DGPs based on the above model (A.1):

$$\begin{aligned} \text{AR}(1): & Y_t = 0.3Y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0,1), \\ \text{AR}(2): & Y_t = 0.3Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0,1). \end{aligned}$$

where the AR(1) is the null model and the AR(2) is the alternative model. For these experiments, we consider $d = 15$. For each replication, we fit it by model (A.1) and obtain the model residual $\widehat{e}_t$ by $\widehat{e}_t = Y_t - \widehat{a}_{0n} - \widehat{a}_{1n} Y_{t-1}$, and use the least squared methods to estimate the parameters $\widehat{a}_{0n}$ and $\widehat{a}_{1n}$. On the basis of the estimation, we calculate our data-driven MDDM-based test statistic $\widehat{AT}_{wn}^{F}$ for different choices of k.

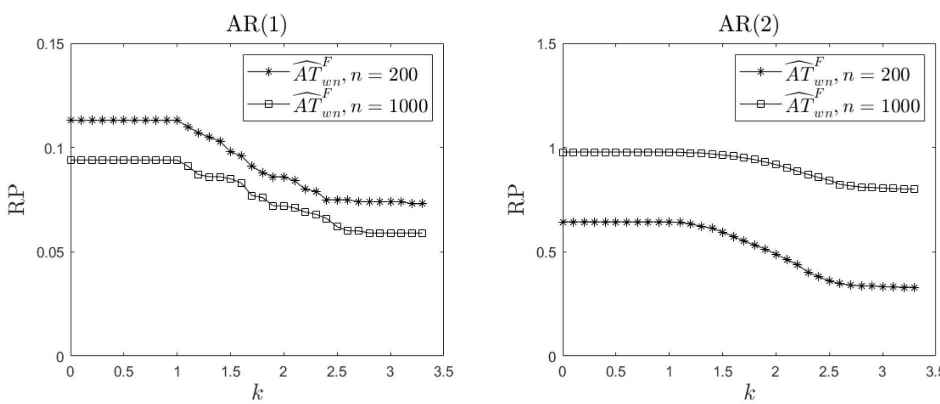

Figure 1: Rejection percentages (5% nominal level) of the tests $\widehat{AT}_{wn}^{F}$ for the models AR(1) and AR(2) for several selected values of tuning parameter k.

Fig.1 shows the empirical rejection percentage (RP) of our proposed test at the 5% level for $k = 0, 0.1, \ldots, 3.3$. The left side of Fig.1 is for the size study, which show that the slope of the empirical RP becomes roughly flat when the values of k exceed 1.8. This indicates that it is not necessary to use a value of k greater than 1.8 in order to properly control the type-I error. The power study are presented in the right side of Fig.1. From the power study, we know that the power is flat when the values of k are between 0 and 1.5, and decreases as k increases. When the values of k exceed 2.5, the slope of the empirical RP plot becomes roughly flat again. On the basis of this analysis, we ultimately select $k = 1.8$ for $p = 1$ as the same for $p = 2, 5$.

A.2 Finite sample tests comparison

In this subsection, we compare the proposed data-driven test $\widehat{AT}_{wn}^{F}$ with the MDDM-based tests $\widehat{T}_{wn}^{F}(M)$ and $\widehat{T}_{sn}^{F}$ proposed by Wang, Zhu and Shao (2022), and the tests $D_{n,C}^2$ and $D_{n,I}^2$ proposed by Escanciano (2006). The test statistics for $D_{n,C}^2$ and $D_{n,I}^2$ are defined as

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

where $n_j = n - j + 1$. The critical values of $D_{n,C}^2$ and $D_{n,I}^2$ are also obtained by the wild bootstrap procedure.

For $p = 1$, we use the same date generation process as in Hong and Lee (2005) to study the finite sample performance. The null model is a univariate AR(1) model :

$$Y_t = a_0 + a_1 Y_{t-1} + \varepsilon_t, \tag{A.2}$$

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 $n = 200, 1000$ from the following two DGPs based on the above model (A.2):

$$\text{DGP 17} : \phi_1 = 1 \text{ and } \phi_2 = 0;$$
$$\text{DGP 18} : \phi_1 = 0.43 \text{ and } \phi_2 = 0.57,$$

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 consider the following eight DGPs:

$$\text{DGP 19}: Y_t = 0.5Y_{t-1} + 0.6Y_{t-1}\varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP 20}: Y_t = 0.5Y_{t-1} - 0.6\varepsilon_{t-1}^2 + \varepsilon_t;$$

$$\text{DGP 21}: Y_t = 0.5Y_{t-1} + 10Y_{t-1}\exp(-Y_{t-1}^2) + \varepsilon_t;$$

$$\text{DGP 22}: Y_t = 0.5Y_{t-1}I(Y_{t-1} \leq 0) - 0.5Y_{t-1}I(Y_{t-1} > 0) + \varepsilon_t;$$

$$\text{DGP 23}: Y_t = 1 - 0.5Y_{t-1} - (4 + 0.4Y_{t-1})/(1 + \exp(-Y_{t-1})) + \varepsilon_t;$$

$$\text{DGP 24}: Y_t = 0.5Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t;$$

$$\text{DGP 25}: Y_t = 0.5Y_{t-1} + \sum_{j=1}^{5} 0.5^j \varepsilon_{t-j}^2 + \varepsilon_t;$$

$$\text{DGP 26}: Y_t = I(Y_{t-6} > 0) - I(Y_{t-6} < 0) + \varepsilon_t,$$

where $\varepsilon_t = \eta_t$.

Table 1: The size and power ($\times 100$) of all tests for DGPs 17–26 at level 5%.

Test	DGP 17		DGP 18		DGP 19		DGP 20		DGP 21	
	n									
	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000
$\widehat{AT}^F_{wn}$	8.7	6.8	8.3	6.8	65.5	73.5	96.4	100.0	100.0	100.0
$\widehat{T}^F_{sn}(3)$	7.7	6.1	7.6	7.0	41.6	61.1	50.7	96.2	100.0	100.0
$\widehat{T}^F_{sn}(6)$	8.6	5.9	8.2	6.4	25.4	38.0	15.0	40.2	99.5	100.0
$\widehat{T}^F_{sn}(9)$	7.8	4.9	7.5	6.4	19.8	28.9	10.1	17.7	97.6	100.0
$\widehat{T}^F_{wn}(3)$	8.3	6.3	7.7	6.8	48.2	65.7	60.2	97.9	100	100.0
$\widehat{T}^F_{wn}(6)$	7.5	5.3	7.7	6.8	34.0	54.1	36.4	84.7	99.9	100.0
$\widehat{T}^F_{wn}(9)$	6.4	4.6	7.4	7.2	28.4	48.4	26.4	70.9	99.5	100.0
$D^2_{n,C}$	6.2	5.4	6.0	6.9	73.6	95.0	97.3	100	100	100
$D^2_{n,I}$	6.2	5.7	6.3	7.4	55.0	91.9	97.3	100	100	100

Test	DGP 17		DGP 18		DGP 19		DGP 20		DGP 21	
	n									
	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000
$\widehat{AT}^F_{wn}$	93.3	100	100	100	95.6	100	85.5	100	100	100
$\widehat{T}^F_{sn}(3)$	26.7	73.7	100	100	93.0	100	53.7	98.4	41.2	51.0
$\widehat{T}^F_{sn}(6)$	12.4	26.7	100	100	75.0	99.4	22.9	53.6	100	100
$\widehat{T}^F_{sn}(9)$	10.2	16.4	99.7	100	60.0	98.1	14.7	27.5	100	100
$\widehat{T}^F_{wn}(3)$	47.3	94.8	100	100	89.5	100	45.5	92.4	41.4	49.3
$\widehat{T}^F_{wn}(6)$	30.9	79.1	100	100	68.5	98.9	27.7	70.3	100	100
$\widehat{T}^F_{wn}(9)$	24.6	65.7	99.9	100	57.7	95.6	22.1	56.0	100	100
$D^2_{n,C}$	91.7	100	100	100	25.0	74.5	78.1	100	69.5	100
$D^2_{n,I}$	84.3	99.7	100	100	80.0	99.9	73.7	100	97.2	100

Table 1 reports the size and power for all examined tests. For these experiments, we consider the upper bound $d = 15$ and the tuning parameter $k = 1.8$. From Table 1, we have the following finds for the size studies:

(1) The proposed data-driven tests $\widehat{AT}^F_{wn}$ has satisfied size performance for most of case, especially for the large sample size $n = 1000$. In general, our proposed methods tend to oversized especially in DGP 18.

(2) Most of the tests tend to oversize at most of cases, especially for the small sample size $n = 200$, while the case is relieved as the sample size increasing. Our proposed data-driven is a little more oversize than the other methods when sample size is $n = 200$, and $\widehat{AT}^F_{wn}$ is similar to the other tests when sample size is $n = 1000$.

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

(1) For all of the DGPs 19-26, our proposed data-driven test $\widehat{AT}^F_{wn}$ has a considerable power. The $\widehat{AT}^F_{wn}$ test has the worst performance for the DGP 19 and the power is 0.655 and 0.735 for $n = 200$ and 1000, respectively. And for the DGPs 20-26, the $\widehat{AT}^F_{wn}$ test usually has power 1.0 for $n = 1000$ and has power more than 0.855 for $n = 200$.

(2) For the MDDM-based methods $\widehat{T}^F_{sn}(M)$, $\widehat{T}^F_{wn}(M)$, the power is decreasing with the lag increasing, i.e. $\widehat{T}^F_{sn}(3)$ and $\widehat{T}^F_{wn}(3)$ have the largest power among $M = 3, 6, 9$, while $\widehat{T}^F_{sn}(9)$ and $\widehat{T}^F_{wn}(9)$ have the worst performance among $M = 3, 6, 9$. Our proposed data-driven test usually performs better than the MDDM-based methods $\widehat{T}^F_{sn}(M)$, $\widehat{T}^F_{wn}(M)$ for all of the considered DGPs 19-26. Our data-driven method has similar performance as $D^2_{n,C}$ for DGPs 20-23, and has better performance than $D^2_{n,C}$ for DGPs 24-26 when $n = 200$, and performs worse than $D^2_{n,C}$ for DGP 19. Our data-driven method has similar performance as $D^2_{n,I}$ for DGPs 20–23 and 26, and has better performance than $D^2_{n,I}$ for DGPs 24-25 when $n = 200$, and performs worse than $D^2_{n,C}$ for DGP 19 when $n = 1000$.

A.3 Simulation for a high-order dependent: AR(10)

To demonstrate our proposed data-driven test has good performance in a high-order dependent. In this subsection, we consider the VAR(10) model:

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

For these experiments, we consider the sample size $n = 1000$, the upper bound $d = 15$ and the tuning parameter $k = 1.8$. Table 2 reports the empirical RP for six values of $\beta = -0.4, -0.3, -0.2, 0.2, 0.3$ and 0.4.

Table 2: Empirical power (percentages) for AR(10) with $n = 1000$.

β	-0.4	-0.3	-0.2	0.2	0.3	0.4
$\widehat{AT}^F_{wn}$	100	91.4	28.0	27.5	91.8	100
$\widehat{T}^F_{sn}(3)$	13.0	10.0	7.9	8.3	11.1	15.1
$\widehat{T}^F_{sn}(6)$	12.2	8.9	6.2	11.4	18.6	29.3
$\widehat{T}^F_{sn}(9)$	64.4	36.1	16.4	21.6	42.3	68.4
$\widehat{T}^F_{wn}(3)$	12.0	8.4	7.1	8.1	10.7	14.0
$\widehat{T}^F_{wn}(6)$	11.8	8.0	7.1	11.7	18.1	27.1
$\widehat{T}^F_{wn}(9)$	62.2	33.7	15.9	20.7	39.8	64.7
$D^2_{n,C}$	15.1	10.8	7.7	7.7	13.4	23.1
$D^2_{n,I}$	42.4	19.5	10.8	11.8	28.0	51.9

From Table 2, 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 is increasing as the absolute value of β increases. For all of the considered six cases, the power of the MDDM-based tests $\widehat{T}^F_{sn}(M)$ and $\widehat{T}^F_{wn}(M)$ are increase as the M increases.

(2) Our proposed data-driven test $\widehat{AT}^F_{wn}$ has a consider power, which is much powerful than all of the other compared tests for all of the six considered values of β. The MDDM-based tests $\widehat{T}^F_{sn}(9)$ and $\widehat{T}^F_{wn}(9)$ have similar performance as the $D^2_{n,I}$ test, and much better than the $\widehat{T}^F_{sn}(3)$, $\widehat{T}^F_{sn}(3)$, $\widehat{T}^F_{wn}(6)$, $\widehat{T}^F_{wn}(6)$ and $D^2_{n,C}$ tests.

A.4 Selection of d

In this subsection, we examine the sensitivity of the proposed data-driven test $\widehat{AT}^F_{wn}$ to the selection of the upper bound d. Similarly, we use the same the null and alternative models in Subsection 1.1. The results are reported in Table 3.

Table 3: RP (percentages) of the data-driven test for different values of d whith nominal 0.05 and $n = 1000$.

d	15	20	25	30	35	40
AR(1)	5.6	5.6	5.6	5.6	5.6	5.6
AR(2)	93.5	93.2	93.2	93.3	93.3	93.8

Table 3 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 have performed additional experiments under the null and under the alternative, for a variety of sample sizes and model specifications, and in all cases we have found the absolute lack of sensitivity to the selection of d.

B. Finite sample performance when moment conditions fail

In this section, we consider the robust of proposed data-driven method proposed in the presented paper. We use the same DPGs 1–16 as in Subsection 5.2 for $p = 2, 5$ but with independent p-dimension $t(4)$ distribution replacing the independent p-dimension normal distribution for the η_t. The Assumptions 1 need an finite 4th moment for the ε_t, the moment conditions fail if η_t is an i.i.d. $t(4)$ distribution. The results are summarized in Tables 4–5.

From Tables 4–5, we have the following finds for the size study:

1) Generally speaking, we can get similar results for the data-driven MDDM-based testing method for the independent $t(4)$ distribution as the independent standard normal distribution.

2) Specifically, for the homoscedasticity error (e.g. DGPs 1 and 9), the difference between the independent $t(4)$ and independent standard normal distribution is very little and can be ignored for both of considered sample size $n = 200, 1000$. While for the heteroscedasticity error (e.g.DGPs 2 and 10), the size is a little larger for the independent $t(4)$ distribution than for the independent standard normal distribution. And this phenomenon is particularly obviously for $p = 5$. We can get similar conclusion for the other considered testing methods.

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

1) Generally speaking, for all of the considered testing methods, we can get similar tendency for the independent $t(4)$ distribution as for the independent standard normal distribution, e.g., the data-driven MDDM-based tests have satisfying power and have better performance than the other considered testing methods and so on.

2) For $p = 2$, the difference can be ignored with replacing the normal distribution by the $t(4)$ distribution. While,for the most cases of $p = 5$, the power is much smaller for the independent $t(4)$ than for the independent standard normal distribution, especially for the small sample size $n = 200$. And for $n = 1000$, this phenomenon is relieved. And we can get similar conclusion for the other considered testing methods.

Table 4: The size and power ($\times 100$) of all tests for DGPs 1–8 at level 5%.

Test	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
	n															
	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000
$\widehat{AT}_{wn}^{F}$	6.8	5.8	7.7	6.2	99.7	100.0	90.0	99.5	99.7	100	100	100	99.6	100.0	100.0	100.0
$\widehat{AT}_{wn}^{S}$	6.6	5.3	7.7	6.1	99.4	100.0	89.5	99.5	99.7	100	100	100	99.5	100.0	100.0	100.0
$\widehat{T}_{sn}^{F}(3)$	6.3	6.0	7.5	6.2	99.8	100.0	82.3	99.6	99.7	100.0	100.0	100.0	97.8	100.0	83.2	100.0
$\widehat{T}_{sn}^{F}(6)$	6.8	7.6	7.7	6.1	94.1	99.9	58.3	96.4	99.1	100	54.9	99.8	63.3	98.6	45.3	92.8
$\widehat{T}_{sn}^{F}(9)$	6.2	6.2	8.0	6.4	80.4	99.8	54.3	90.9	96.5	100	27.2	72.9	41.1	93.3	31.7	74.7
$\widehat{T}_{sn}^{S}(3)$	6.2	5.8	7.4	6.4	99.3	100	81.8	99.4	99.6	100.0	99.7	100.0	97.3	100.0	80.1	100.0
$\widehat{T}_{sn}^{S}(6)$	6.6	7.2	7.3	5.9	91.4	99.9	58.0	96.2	98.8	100.0	39.8	96.6	63.6	98.7	43.0	91.2
$\widehat{T}_{sn}^{S}(9)$	6.6	6.2	8.0	5.9	77.1	99.7	54.3	90.9	95.4	100.0	20.2	55.6	39.6	93.2	29.8	71.6
$\widehat{T}_{wn}^{F}(3)$	6.6	7.7	8.1	6.9	99.4	100.0	78.8	99.4	99.8	100.0	100.0	100.0	99.2	100.0	96.6	100.0
$\widehat{T}_{wn}^{F}(6)$	6.6	7.0	9.2	7.0	92.4	99.9	57.3	94.9	99.5	100.0	100.0	100.0	81.9	100	84.4	100.0
$\widehat{T}_{wn}^{F}(9)$	6.7	5.7	8.2	7.0	77.5	99.8	52.7	89.0	98.2	100.0	100.0	100.0	66.3	99.7	71.7	99.7
$\widehat{T}_{wn}^{S}(3)$	6.6	7.6	8.3	6.8	99.4	100	77.8	99.2	99.6	100.0	100.0	100.0	99.1	100	96.0	100.0
$\widehat{T}_{wn}^{S}(6)$	6.7	6.9	8.4	6.7	91.3	99.9	56.5	94.7	99.4	100.0	100.0	100.0	81.4	100.0	82.0	100.0
$\widehat{T}_{wn}^{S}(9)$	6.4	5.9	7.9	6.9	75.2	99.8	51.7	89.2	97.3	100.0	100.0	100.0	65.3	99.6	70.1	99.6
$\widehat{Q}_1(3)$	5.6	6.0	9.1	10.8	100.0	100.0	100.0	100.0	98.2	100.0	5.0	4.8	62.1	93.9	21.7	39.9
$\widehat{Q}_1(6)$	5.5	6.0	10.8	10.3	100.0	100.0	100.0	100.0	96.2	100.0	7.1	5.5	53.8	92.7	15.8	28.8
$\widehat{Q}_1(9)$	6.4	4.9	10.7	9.6	100.0	100.0	99.9	100.0	95.2	100.0	7.0	4.2	45.8	87.0	14.5	25.9
$\widehat{Q}_2(3)$	5.5	6.1	9.6	11.1	100.0	100.0	100.0	100.0	98.2	100.0	5.4	4.5	62.6	94.2	22.2	40.5
$\widehat{Q}_2(6)$	5.3	6.2	11.9	10.7	100.0	100.0	100.0	100.0	96.7	100.0	7.8	5.7	56.7	93.1	17.3	29.7
$\widehat{Q}_2(9)$	5.8	5.6	12.2	10.5	100.0	100.0	99.9	100.0	95.9	100.0	8.4	5.0	49.7	87.7	16.6	27.0
$\widehat{Q}_3(3)$	3.2	6.1	9.3	10.9	100.0	100.0	100.0	100.0	98.2	100.0	5.3	4.5	62.6	94.2	22.2	40.0
$\widehat{Q}_3(6)$	4.0	6.2	11.6	10.6	100.0	100.0	100.0	100.0	96.6	100.0	7.5	5.7	56.0	93.1	17.0	29.5
$\widehat{Q}_3(9)$	3.8	5.3	11.8	10.1	100.0	100.0	99.9	100.0	95.9	100.0	7.8	4.8	49.2	97.6	16.1	26.6
$\widehat{LM}(3)$	4.9	5.7	9.8	10.6	100.0	100.0	100.0	100.0	99.1	100.0	2.2	2.3	63.0	97.0	13.8	29.9
$\widehat{LM}(6)$	4.2	3.9	9.1	19.0	100.0	100.0	100.0	100.0	97.9	100.0	3.6	2.1	46.7	90.6	12.6	23.2
$\widehat{LM}(9)$	3.2	5.1	7.3	8.6	100.0	100.0	100.0	100.0	96.7	100.0	2.8	2.1	34.5	83.7	9.0	21.8

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

Test	DGP 9		DGP 10		DGP 11		DGP 12		DGP 13		DGP 14		DGP 15		DGP 16	
	n															
	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000	200	1,000
$\widehat{AT}_{wn}^{F}$	9.8	7.6	8.7	7.3	73.9	99.6	56.7	87.8	100	100.0	100.0	100.0	46.2	89.2	55.3	89.8
$\widehat{AT}_{wn}^{S}$	8.2	6.7	8.6	7.4	42.9	98.4	27.9	43.0	99.7	100.0	100.0	100.0	21.6	41.3	34.6	58.5
$\widehat{T}_{sn}^{F}(3)$	11.6	7.9	8.7	7.5	85.2	99.9	62.1	94.2	100.0	100.0	69.1	100.0	53.4	94.6	13.4	15.0
$\widehat{T}_{sn}^{F}(6)$	11.2	9.1	8.5	8.1	72.0	99.0	50.7	80.6	99.9	100.0	28.6	60.7	40.8	80.5	13.7	11.6
$\widehat{T}_{sn}^{F}(9)$	12.8	10.7	10.2	7.8	66.0	97.0	49.0	72.5	99.4	100	23.9	32.7	39.2	70.5	13.7	11.8
$\widehat{T}_{sn}^{S}(3)$	9.7	6.8	8.3	7.4	64.6	98.0	47.2	75.0	99.8	100.0	35.7	96.0	35.9	79.0	12.0	11.1
$\widehat{T}_{sn}^{S}(6)$	9.5	7.7	8.3	8.0	48.1	89.4	32.2	52.7	99.4	100.0	14.8	30.6	25.4	54.8	8.8	10.6
$\widehat{T}_{sn}^{S}(9)$	9.7	9.2	9.7	7.7	66.0	77.8	29.9	96.9	96.9	100.0	13.8	17.3	21.9	44.8	10.3	11.0
$\widehat{T}_{wn}^{F}(3)$	12.2	6.9	9.0	8.1	79.9	99.9	60.1	91.8	100	100.0	99.9	100.0	56.0	95.7	16.6	21.3
$\widehat{T}_{wn}^{F}(6)$	10.8	7.3	9.6	7.7	65.3	97.5	49.0	74.9	99.9	100.0	94.2	100.0	42.0	82.8	16.3	17.0
$\widehat{T}_{wn}^{F}(9)$	12.4	9.5	10.9	6.9	59.9	92.4	48.5	67.4	99.7	100.0	83.7	100.0	40.7	73.7	17.5	17.3
$\widehat{T}_{wn}^{S}(3)$	10.8	6.8	8.8	7.7	63.8	98.0	45.8	78.6	99.8	100.0	95.7	100.0	41.0	83.9	13.8	16.9
$\widehat{T}_{wn}^{S}(6)$	11.3	6.7	9.3	7.6	51.3	88.5	36.9	57.7	99.7	100.0	74.2	100.0	32.0	67.1	11.5	14.8
$\widehat{T}_{wn}^{S}(9)$	11.8	7.8	10.8	6.5	44.7	80.4	37.9	48.4	98.9	100.0	63.7	99.5	31.4	58.4	13.5	12.6
$\widehat{Q}_1(3)$	6.8	5.3	8.0	9.4	83.1	100.0	65.7	99.8	100.0	100.0	6.4	5.5	39.5	90.1	5.3	4.4
$\widehat{Q}_1(6)$	5.2	5.1	8.7	11.4	69.2	99.9	51.4	95.6	99.7	100.0	6.4	6.4	30.5	76.5	6.2	4.9
$\widehat{Q}_1(9)$	7.7	4.9	10.4	10.6	67.3	99.3	48.9	91.0	99.3	100.0	8.3	6.6	35.4	67.1	6.8	5.5
$\widehat{Q}_2(3)$	7.4	5.5	9.1	9.9	84.7	100.0	68.7	99.8	100.0	100.0	7.3	5.8	41.9	90.7	6.5	4.7
$\widehat{Q}_2(6)$	7.7	6.1	12.1	12.6	73.2	99.9	56.9	96.2	99.9	100.0	8.8	7.2	35.8	78.5	7.9	5.3
$\widehat{Q}_2(9)$	12.2	6.6	16.2	11.8	75.0	99.5	59.7	92.1	99.6	100.0	12.9	8.3	40.6	71.6	11.8	7.1
$\widehat{Q}_3(3)$	7.4	5.5	9.1	9.9	84.7	100.0	68.7	99.8	100.0	100.0	7.2	5.8	41.8	90.7	6.4	4.7
$\widehat{Q}_3(6)$	7.3	6.0	11.9	12.5	73.0	99.9	56.5	96.3	99.9	100.0	8.1	7.0	35.2	78.5	7.7	5.2
$\widehat{Q}_3(9)$	11.2	6.4	15.7	11.6	74.5	99.5	58.9	92.1	99.6	100.0	12.3	8.1	40.0	71.4	11.2	7.0
$\widehat{LM}(3)$	4.6	4.7	6.4	9.4	67.4	100	57.5	99.2	100.0	100.0	2.4	1.2	23.8	83.3	3.2	4.6
$\widehat{LM}(6)$	4.4	3.7	5.6	6.7	41.3	98.2	30.8	91.6	99.8	100.0	3.2	3.2	14.3	60.1	4.9	5.8
$\widehat{LM}(9)$	5.1	4.6	4.7	5.2	28.1	96.7	23.1	78.9	99.3	100.0	2.8	3.4	12.1	47.9	6.0	6.1

C. Proofs

C.1 Proof of Lemma 1

Following Shao and Zhang (2014), we rewrite $||\text{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F$ as

$$||\text{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F = \left\| \frac{1}{c_p} \int_{\mathcal{R}^p} \frac{\widehat{\mathcal{G}}_n^j(s)\widehat{\mathcal{G}}_n^j(s)^\star}{\|s\|^{1+p}} ds \right\|_F, \tag{C.3}$$

where

$$\widehat{\mathcal{G}}_n^j(s) = \frac{1}{n}\sum_{t=1}^n \widehat{e}_t e^{i\langle s, Y_{t-j}\rangle} - \left(\frac{1}{n}\sum_{t=1}^n \widehat{e}_t\right)\left(\frac{1}{n}\sum_{t=1}^n e^{i\langle s, Y_{t-j}\rangle}\right). \tag{C.4}$$

Let $\Omega \subset \mathcal{R}^p$ is a compact set. We know $\sqrt{n-j}\widehat{\mathcal{G}}_n^j(s) \Rightarrow \chi^j(s)$ by the Theorem 3.1 in Wang, Zhu and Shao (2022), where "$\Rightarrow$" denotes weak convergence in $\mathcal{C}$ and $\mathcal{C}$ represents the space of continuous complex-valued random functions over Ω equipped with uniform topology. By the continuous mapping theorem, we can get the conclusion of Lemma 1.

C.2 Proof of Theorem 1

Define

$$M_{BIC} = \min\{M : 1 \le M \le d; L_{BIC,M} \ge L_{BIC,h}, h = 1, 2, \cdots, d\},$$

where

$$L_{BIC,M} = n||\text{MDDM}_n(\widehat{e}_t|Y_{t-M})||_F - \log(3p)\cdot M \log n.$$

Under the null hypothesis and the Assumptions 1-5, we need to prove that,

$$\lim_{n\to\infty} P(M^* = M_{BIC}) = 1, \tag{C.5}$$

and that

$$\lim_{n\to\infty} P(M_{BIC} = 1) = 1. \tag{C.6}$$

To proof (C.5), we define the following events

$$A_n(k) = \left\{ \max_{1\le j\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F > \log(3p)^k\sqrt{\log n} \right\}.$$

By Lamma 1, we can get $\max_{1\le j\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F = O_p(1)$, it follows that

$$P(A_n(k)) = P\left(\max_{1\le j\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F > \log(3p)^k\sqrt{\log n}\right) = 0, \quad as \quad n \to \infty.$$

Then,

$$P((A_n(k))^c) = 1, \quad as \quad n \to \infty,$$

where $(A_n(k))^c$ denotes the complementary set of $A_n(k)$. By the definition of $\pi(M, n, k)$, we know that (C.5) holds.

In the following, we will prove (C.6). Notice that

$$P(M_{BIC} = 1) = 1 - \sum_{j=2}^{d} P(M_{BIC} = j) \ge 1 - \sum_{j=2}^{d} P(L_{BIC,j} \ge L_{BIC,1}). \qquad \text{(C.7)}$$

Now, for $1 \le j \le d$,

$$\begin{aligned} P(L_{BIC,j} \ge L_{BIC,1}) &= P(n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F - \log(3p)(j\log n) \\ &\qquad \ge n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-1})||_F - \log(3p)(\log n)) \\ &\le P(n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-j})||_F \ge (j-1)\log(3p)\log n) \\ &\le P\left(n\sum_{i=1}^{j}||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-i})||_F \ge (j-1)\log(3p)\log n\right) \\ &\le \sum_{i=1}^{j} P\left(n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-i})||_F \ge \frac{j-1}{j}\log(3p)\log n\right). \end{aligned}$$

Since under the null $\max_{1\le i\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-i})||_F = O_p(1)$, it is prove that

$$P(n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-i})||_F \ge (j-1)/j \cdot \log(3p)\log n) \to 0, \text{ as n} \to \infty.$$

Therefore, (C.6) holds, and Theorem 1 follows from an application of the result of

Lemma 1.

C.3 Proof of Theorem 2

Similar as prove the Theorem 1, we define

$$M_{AIC} = \min\{M : 1 \le M \le d; L_{AIC,M} \ge L_{AIC,h}, h = 1, 2, \cdots, d\},$$

where

$$L_{AIC,M} = n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-M})||_F - \log(3p) \cdot 2M.$$

Under the alternative H_1^K and the conditions for Theorem 2 hold, we will prove that

$$\lim_{n\to\infty} P(M^* = M_{AIC}) = 1, \tag{C.8}$$

and that

$$P(M_{AIC} \ge K) \to 1. \tag{C.9}$$

We define the event

$$B_n(k) = \left\{ \max_{1\le i\le d} n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-i})||_F < \log(3p)^k \sqrt{\log n} \right\}.$$

Then, since $K \le d$,

$$P(B_n(k)) \le P\left(n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-K})||_F < \log(3p)^k \sqrt{\log n} \right) \to 0.$$

By the theorem 3.2 of Wang, Zhu and Shao (2022) and the alternative H_1^K, we know that,

$$||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-K})||_F \to_p ||\mathrm{MDDM}(\varepsilon_t|Y_{t-K})||_F \ne 0.$$

Then, by the definition of $\pi(M, n, k)$,we know that (C.8) holds.

Now, for $k^* = 1, \cdots, K-1$,

$$\begin{aligned} P(M_{AIC} = k^*) &\le P(L_{AIC,k^*} \ge L_{AIC,K}) \\ &= P(n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-k^*})||_F - \log(3p) \cdot 2k^* \end{aligned}$$

$$\begin{aligned} &\geq n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-K})||_F - \log(3p)\cdot 2K) \\ = P(n&||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-k^*})||_F \\ &\geq 2(k^*-K)\log(3p) + n||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-K})||_F) \to 0. \end{aligned}$$

Since

$$||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-k^*})||_F \to 0 \text{ and } ||\mathrm{MDDM}_n(\widehat{e}_t|Y_{t-K})||_F \to ||\mathrm{MDDM}(\varepsilon_t|Y_{t-K})||_F \neq 0.$$

Hence, (C.9) holds. Therefore, for each $C>0$,

$$\begin{aligned} P(\widehat{T}^F_{wn}(M^*) \leq C) &= P(\widehat{T}^F_{wn}(M^*) \leq C, M^* \geq K) + o(1) \\ &\leq P\left((n-K)||\mathrm{MDDM}_n(\hat{e}_t|Y_{t-K})||_F \leq C\right) + o(1) \\ &= o(1). \end{aligned}$$

Then, $A\widehat{T}^F_{wn} \to \infty$ as $n\to\infty$, and the test statistic is consistent against H_1^K.

C.4 Proof of Lemma 2

Similar to (C.3), we have

$$\|\mathrm{MDDM}_n(\widehat{e}_t^{**}|Y_{t-j})\|_F = \left\| \frac{1}{c_p}\int_{\mathcal{R}^p} \frac{\widehat{\mathcal{G}}_n^{j*}(s)\widehat{\mathcal{G}}_n^{j*}(s)^\star}{\|s\|^{1+p}} ds \right\|_F, \tag{C.10}$$

where $\widehat{\mathcal{G}}_n^{j*}(s)$ is defined in the same way as $\widehat{\mathcal{G}}_n^{j}(s)$ in (C.4) with $\widehat{e}_t^{**}$ replaced by $\widehat{e}_t^{*}$. By the Theorem 4.1 in Wang et al. (2022), we have

$$\sqrt{n}\widehat{\mathcal{G}}_n^{j*}(s) \Rightarrow \chi_*^j(s) \text{ in probability.}$$

By the Corollary 4.1 in Wang, Zhu and Shao (2022), we have

$$(n-j)\|\mathrm{MDDM}_n(\widehat{e}_t^{**}|Y_{t-j})\|_F \to_d \left\| \frac{1}{c_p}\int_{\mathcal{R}^k} \frac{\chi_*^j(s)\chi_*^j(s)^\star}{\|s\|^{1+p}} ds \right\|_F \text{ in probability.}$$

Therefore,

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

$$
\begin{aligned}
&= n\sum_{j=1}^{M^*}\omega_j\|\mathrm{MDDM}_n(\widehat{e}_t^{**}|Y_{t-j})\|_F \\
&= \sum_{j=1}^{M^*}(n-j)\|\mathrm{MDDM}_n(\widehat{e}_t^{**}|Y_{t-j})\|_F \\
&\to_d \sum_{j=1}^{M^*}\left\|\frac{1}{c_p}\int_{\mathcal{R}^p}\frac{\chi_*^j(s)\chi_*^j(s)^\star}{\|s\|^{1+p}}ds\right\|_F \text{ in probability.}
\end{aligned}
$$

Hence, Lemma 2 holds.

References

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.

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.

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.

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.