Sign-based specification tests for martingale difference with conditional heteroscedasity

Chen, Min and Zhu, Ke

Chinese Academy of Sciences and Capital University of Economics and Business, Chinese Academy of Sciences

1 June 2014

Sign-based specification tests for martingale difference with conditional heteroscedasity

BY MIN CHEN

Chinese Academy of Sciences and Capital University of Economics and Business

mchen@amss.ac.cn

AND KE ZHU

Chinese Academy of Sciences

kzhu@amss.ac.cn

ABSTRACT

This article proposes Cramér-von Mises (CM) and Kolmogrove-Smirnov (KS) test statistics based on the signs of a time series to test the null hypothesis that the series is a martingale difference sequence (MDS) with conditional heteroscedasity. Both of test statistics allowing for heavy-tailedness, non-stationarity, and nonlinear serial dependence of unknown forms, are easy-to-implement. Unlike the sign-based variance-ratio test in Wright (2000), our sign-based CM and KS tests have no need to select the lag. Unlike other often used specification tests for MDS, our sign-based CM and KS tests are robust and have exact distributions which can be simulated easily. Simulation studies and applications further demonstrate the importance of our sign-based CM and KS tests.

Some key words: Conditional heteroscedasity; Cramér-von Mises test; Kolmogrove-Smirnov test; Martingale difference; Robustness.

1. INTRODUCTION

One of the most important questions in applied econometrics and empirical finance is the issue of whether a time series such as the stock-market and exchange-rate returns forms a martingale difference sequence (MDS). This issue is closely related to market efficiency in the weak form;

see, e.g., Timmermann and Granger (2004) and Lim and Brooks (2011). Once a time series is a MDS, it is unpredictable; otherwise, there has a practical demand to fit its conditional mean by some useful models. Thus, testing for MDS is meaningful and has been popular in the literature; see, e.g., Durlauf (1991) and Deo (2000) for earlier works and Escanciano (2007) and the references therein for more recent ones. Moreover, for most of economics and financial data y_t, if it is a MDS, it often admits a conditional heteroscedastic form as follows:

$$y_t = \varepsilon_t \sigma_t, \tag{1.1}$$

where $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $\sigma_t \in \mathcal{F}_{t-1}$ is positive, and $\mathcal{F}_t$ is a σ-field generated by $\{y_t, y_{t-1}, \cdots\}$. This feature of y_t has been more or less accepted in application after the seminal work of Engle (1982) and Bollerslev (1986). Many existing models, such as the GARCH model and its vast variants (see, e.g., Fan and Yao (2003) or Tsay (2005) for an overview), the stochastic volatility model in Shephard (1996), the conditional piecewise constant volatility model in Chan et al. (2014) to name but a few, are nested by model (1.1). Thus, it is desirable to detect the following null hypothesis:

$$H_0 : y_t \text{ admits the form as in } (1.1).$$

Needless to say, the Cramér-von Mises (CM) test based on sample autocorrelations of $\{y_t\}$ in Deo (2000) and the CM and Kolmogrove-Smirnov (KS) tests based on some marked processes in Escanciano (2007) are both valid for this purpose; see also Hong (1996, 1999), Shao (2011a, b), Zhu and Li (2014) and references therein for testing the null hypothesis that y_t is a white noise. However, when y_t is heavy-tailed with an infinite variance (see, e.g., Davis and Mikosch (1998), Rachev (2003) and Zhu and Ling (2014) for some empirical examples in this context), none of existing tests except the sign-based variance-ratio (VR) test in Wright (2000) is feasible. In this paper, we propose the sign-based CM and KS tests to detect H_0. Both of our sign-based CM and KS tests allowing for heavy-tailedness, non-stationarity, and nonlinear serial dependence of unknown forms, are easy-to-implement. Unlike the sign-based VR test in Wright (2000), our sign-based CM and KS tests have no need to select the lag. Unlike other aforementioned tests for MDS, our sign-based CM and KS tests are robust and have exact distributions which can be simulated easily. Simulation studies and applications further demonstrate the importance of our tests.

This paper is organized as follows. Section 2 gives our test statistics and their exact distributions. Simulation results are reported in Section 3. Applications are given in Section 4. Concluding remarks are offered in Section 5.

2. Test statistic and exact distribution

The use of sign-based tests in regression and time series models so far has attracted considerable interest; see, e.g., Koenker and Bassett (1982), Wright (2000), Hallin et al. (2008), Coudin and Dufour (2009), Chen and Zhu (2014), Zhu and Ling (2014), and many others. In this section, based on the signs of $\{y_t\}$, we propose the CM and KS tests to detect H_0.

Denote by $sgn(y_t) := 2I(y_t > 0) - 1$ the sign of y_t, where $I(\cdot)$ is the indicator function. Let $\gamma(j) = cov(s_t, s_{t+j})$ with $s_t = sgn(y_t)$. Then, the spectral density function and spectral distribution function of s_t, respectively, are

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega} \quad \text{for } \omega \in [-\pi, \pi]$$

and $F(\lambda) = \int_0^{\lambda} f(\omega) d\omega$ for $\lambda \in [0, \pi]$. Following Shao (2011a), the sample spectral distribution function of s_t is

$$F_n(\lambda) = \sum_{j=0}^{n-1} \hat{\gamma}(j) \psi_j(\lambda),$$

where $\hat{\gamma}(j) = n^{-1} \sum_{t=1+|j|}^{n} (s_t - \bar{s})(s_{t-|j|} - \bar{s})$ is the sample autocovariance function of s_t at lag j, $\bar{s} = n^{-1} \sum_{t=1}^{n} s_t$ is the sample mean of s_t, and

$$\psi_j(\lambda) = \begin{cases} \sin(j\lambda)/j\pi & \text{if } j \neq 0 \\ \lambda/2\pi & \text{if } j = 0 \end{cases}.$$

Moreover, to validate our test statistics, we need the following assumption as in Wright (2000):

Assumption 2.1. $I(\varepsilon_t > 0)$ is an i.i.d. binomial random variable that is 1 with probability $1/2$ and 0 otherwise.

Assumption 2.1 holds when ε_t is an i.i.d. random variable with a median zero and a continuous pdf at zero, and hence it allows for the heavy-tailed ε_t as in Berkes and Horváth (2004), Linton et al. (2010), and Chen and Zhu (2014). In addition, since the i.i.d. assumption on ε_t is not necessary from Assumption 2.1, ε_t could also be t-distributed with time-varying degrees of freedom as considered by Hansen (1994).

Based on Assumption 2.1, it is straightforward to see that under H_0, $\{s_t\}$ is an i.i.d. binomial variable that is 1 with probability $1/2$ and -1 otherwise. This implies that $F(\lambda) = \gamma(0)\psi_0(\lambda)$ under H_0, and the sample spectral distribution $F_n(\lambda)$ becomes $\hat{\gamma}(0)\psi_0(\lambda)$ in this case. Thus, it is reasonable to consider the following sign-based CM and KS test statistics to detect H_0, where

$$\text{CM}_n = \int_0^{\pi} S_n^2(\lambda)d\lambda, \quad \text{KS}_n = \max_{\lambda\in[0,\pi]} S_n^2(\lambda), \tag{2.1}$$

and the process

$$S_n(\lambda) = \sqrt{n}\left\{F_n(\lambda) - \hat{\gamma}(0)\psi_0(\lambda)\right\} =: \sum_{j=1}^{n-1} \sqrt{n}\hat{\gamma}(j)\psi_j(\lambda)$$

measures the distance between $F_n(\lambda)$ and $\hat{\gamma}(0)\psi_0(\lambda)$. Clearly, CM_n or KS_n takes into account of the autocorrelations of s_t at all lags, and a large value of CM_n or KS_n is in favor of rejecting H_0. Next, we give the exact null distributions of CM_n and KS_n in the following theorem:

THEOREM 2.1. *Suppose that Assumption* 2.1 *holds. Under H_0, CM_n and KS_n have the same distribution as*

$$\int_0^{\pi} \left[S_n^*(\lambda)\right]^2 d\lambda \;\; and \;\; \max_{\lambda\in[0,\pi]} \left[S_n^*(\lambda)\right]^2,$$

respectively, where

$$S_n^*(\lambda) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n-1} \sum_{t=1+j}^{n} (s_t^* - \bar{s}^*)(s_{t-j}^* - \bar{s}^*)\psi_j(\lambda),$$

$\{s_t^*\}_{t=1}^n$ *is an i.i.d. sequence, each element of which is* 1 *with probability* $1/2$ *and* -1 *otherwise, and* $\bar{s}^* = n^{-1}\sum_{t=1}^n s_t^*$.

Remark 2.1. The CM test based on $\{y_t\}$ itself has been studied by Shao (2011a), and our CM test based on $\{s_t\}$ can be viewed as a robust version of his test. Compared to the CM test in Shao (2011a) and the CM and KS tests in Escanciano (2007), our sign-based CM and KS tests only take into account of the signs of $\{y_t\}$, and hence they may not be consistent. But, our sign-based tests also have three potential advantages: first, neither finite second moment condition nor stationarity of y_t is needed; second, no bootstrap procedure is required to obtain the critical values; third, the technical difficulty in proving the tightness for the test statistic is circumvented.

Remark 2.2. From Theorem 2.1, we know that our sign-based CM and KS tests allow for heavy-tailedness, non-stationarity, and nonlinear serial dependence of unknown forms. This is also the case for the Wright's (2000) sign-based variance-ratio (VR) test defined by

$$\mathrm{VR}_n(k) = \left[\frac{\sum_{t=k+1}^{n}(s_t + s_{t-1} + \cdots + s_{t-k+1})^2}{k\sum_{t=1}^{n} s_t^2} - 1\right] \times \left[\frac{2(2k-1)(k-1)}{3kn}\right]^{-1/2},$$

where k is the lag parameter. Numerical studies in Wright (2000) showed that the performance of $\mathrm{VR}_n(k)$ is sensitive to k, but how to choose the optimal k is hard in theory. Our sign-based CM and KS tests do not face this dilemma, since they are free of user-chosen parameter.

Remark 2.3. In application, one may consider the following null hypothesis H_0' instead of H_0:

$$H_0' : y_t = \mu + \varepsilon_t \sigma_t,$$

where μ is an unknown parameter. To detect H_0', it may be natural to first estimate μ by $\mathrm{med}(y_t)$ (i.e., the sample median of $\{y_t\}$), and then apply both CM_n and KS_n to the adjusted series $\{\tilde{y}_t\}$, where $\tilde{y}_t = y_t - \mathrm{med}(y_t)$. However, we can see that $sgn(\tilde{y}_t) \neq sgn(\varepsilon_t)$ due to the estimation of μ. Thus, this aforementioned method is not valid, and the portmanteau test based on the bootstrap method in Zhu and Ling (2014) should be used to detect H_0'.

Based on 20,000 repetitions, Table 1 reports the $100(1-\alpha)\%$ percentiles of the exact null distributions of CM_n and KS_n for some choices of n and α. Then, we reject H_0 at the significance level α, when the value of CM_n or KS_n is larger than the corresponding percentile.

Table 1. $100(1-\alpha)\%$ *percentiles of the exact null distributions of* CM_n *and* KS_n.

		α				
Tests	n	0.900	0.925	0.950	0.975	0.990
CM_n	100	0.5255	0.6016	0.7024	0.8960	1.1612
	500	0.5545	0.6278	0.7316	0.9244	1.1871
	1000	0.5399	0.6151	0.7139	0.8860	1.1395
KS_n	100	0.6087	0.6737	0.7659	0.9367	1.1505
	500	0.6981	0.7645	0.8634	1.0224	1.2447
	1000	0.6976	0.7709	0.8624	1.0333	1.2469

3. SIMULATIONS

There are an enormous number of ways of testing H_0; see, e.g., Chen and Deo (2006), Shao (2011a), and refrences therein. The simulation studies we conduct in this section do not attempt to compare all possible tests but only the sign-based VR test in Wright (2000), because the sign-based VR test enjoys the same simplicity and robustness as ours. The goal is limited. But if our sign-based CM and KS tests have better size and power properties than the sign-based VR test in some plausible models, then it follows that our new tests should be useful specification tests for practitioners.

The models we use to examine the size and power performance of all sign-based tests are as follows:

$$\text{Model 1}: y_t = \varepsilon_t \exp(h_t/2),\ h_t = 0.95h_{t-1} + \xi_t,\ \xi_t \sim \text{i.i.d.}N(0, 0.1),$$
$$\text{Model 2}: y_t = \varepsilon_t\sqrt{h_t},\ h_t = 0.1 + [0.2 + 0.1I(\varepsilon_{t-1} < 0)]y_{t-1}^2 + 0.8h_{t-1},$$
$$\text{Model 3}: y_t = \varepsilon_t\sqrt{h_t},\ h_t = 0.1 + 0.1[1 - 0.4sgn(\varepsilon_{t-1}) + 0.04]y_{t-1}^2 + 0.9h_{t-1},$$
$$\text{Model 4}: y_t = \varepsilon_t\sqrt{h_t},\ h_t = 0.1 + 0.147y_{t-1}^2 + 0.926h_{t-1},$$
$$\text{Model 5}: y_t = 0.1y_{t-1} + \nu_t,\ \text{where } \nu_t \text{ is defined as } y_t \text{ in model } 1,$$
$$\text{Model 6}: y_t = 0.1y_{t-1} + \nu_t,\ \text{where } \nu_t \text{ is defined as } y_t \text{ in model } 2,$$
$$\text{Model 7}: y_t = 0.2y_{t-2} + \nu_t,\ \text{where } \nu_t \text{ is defined as } y_t \text{ in model } 3,$$
$$\text{Model 8}: y_t = 0.2y_{t-2} + \nu_t,\ \text{where } \nu_t \text{ is defined as } y_t \text{ in model } 4.$$

Clearly, models 1-4 which admit the specification of MDS with conditional heteroscedasticity are used for the size simulation study, and the remaining models are used for the power simulation study. Specifically, model (1) is the stochastic volatility model of conditional heteroscedasticity used in Wright (2000); models (2)-(3) used in Zhu and Ling (2014) are GJR model and non-linear GARCH model with $Ey_t^2 = \infty$, respectively; model (4) is a non-stationary GARCH model used in Francq and Zakoïan (2012) for fitting KVA series; and models (5)-(8) deviate from models (1)-(4) by an AR model, respectively. In each of these eight models, ε_t is i.i.d. standard normal, standardized t_3 with variance one, or t_{η_t}, where the degree of freedom η_t is dynamically generated from $\eta_t = 27.9/(1 + \exp(-\lambda_t)) + 2.1$ and $\lambda_t = -1.07 - 0.38\varepsilon_{t-1} - 0.08\varepsilon_{t-1}^2$. The last choice for t_{η_t} is taken from Hansen (1994), who used the logistic transformation to bound the degrees of freedom η_t between 2.1 and 30.

As Wright (2000), we generate 5000 replications of sample size $n = 100$, 500 and 1000 from each aforementioned model. For each replication, we use the tests CM$_n$, KS$_n$, and VR$_n(k)$ for $k = 2, 5, 10$ to detect H_0. Table 2 reports the size and power of all tests based on the significance level $\alpha = 0.05$, where the critical values of CM$_n$ and KS$_n$ are taken from Table 1, and the critical values of VR$_n(k)$ are taken from Table 1 in Wright (2000). From Table 2, it is clear that the sizes of these tests are close to their nominal ones, and the power of them is generally as expected. First, all the powers become large as n increases. Second, each test has a larger power when the tail of ε_t is heavier. Third, when y_t exhibits the first order autocorrelation as in models 5-6, the

Table 2. *Empirical sizes and power ($\times 100$) for all sign-based tests.*

Tests	n	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
		Panel A: ε_t is standard normal							
CM_n	100	4.84	4.80	5.44	5.10	7.84	8.10	10.44	10.08
	500	4.52	4.96	4.88	5.02	27.30	23.02	26.50	22.90
	1000	5.36	4.80	4.84	4.62	52.48	44.00	62.54	53.20
KS_n	100	4.76	4.76	5.06	5.16	7.80	7.90	10.86	10.78
	500	4.88	5.16	4.80	4.72	24.18	21.36	30.96	27.12
	1000	5.38	4.98	4.74	4.70	48.26	40.06	61.94	53.78
$\text{VR}_n(2)$	100	4.32	4.32	4.68	4.38	7.46	6.86	7.46	7.60
	500	4.56	5.06	4.84	4.82	28.00	23.70	8.00	7.96
	1000	5.46	5.28	5.10	4.80	52.54	44.22	9.48	8.88
$\text{VR}_n(5)$	100	4.80	4.66	4.84	4.56	6.86	8.04	10.86	9.46
	500	4.66	5.48	4.76	4.74	19.10	16.36	32.96	29.40
	1000	5.26	4.88	4.66	4.82	34.96	29.16	57.02	51.36
$\text{VR}_n(10)$	100	4.90	5.30	4.98	5.34	7.02	7.48	10.76	8.94
	500	4.64	4.28	4.98	4.94	12.20	10.58	27.64	25.46
	1000	5.44	4.12	4.76	4.68	21.26	17.42	49.90	44.96
		Panel B: ε_t is standardized t_3							
CM_n	100	4.66	4.88	5.28	5.12	10.10	8.64	12.48	11.38
	500	5.38	4.54	4.56	4.98	40.96	31.40	43.02	37.86
	1000	5.56	4.86	5.06	4.92	70.32	60.02	84.60	79.16
KS_n	100	4.96	5.08	5.22	4.82	9.92	8.34	14.58	12.76
	500	5.22	4.64	4.80	4.92	36.72	28.82	45.70	41.16
	1000	5.60	4.66	5.32	5.42	66.00	55.98	80.56	76.76
$\text{VR}_n(2)$	100	4.22	4.34	5.02	4.30	9.20	7.82	8.64	7.32
	500	5.30	4.44	4.50	5.14	41.74	32.22	9.78	9.04
	1000	5.86	5.06	5.24	5.34	71.04	59.82	9.36	9.42
$\text{VR}_n(5)$	100	4.52	4.44	4.50	4.44	7.96	7.74	12.88	11.38
	500	5.12	4.54	5.12	5.30	26.60	22.38	44.60	41.34
	1000	5.08	4.54	5.26	5.22	48.50	39.60	72.02	68.52
$\text{VR}_n(10)$	100	5.28	5.26	4.60	4.62	6.82	6.90	12.20	11.36
	500	4.72	4.76	5.04	4.48	15.94	13.54	38.40	34.50
	1000	5.04	4.76	5.02	4.60	28.72	23.14	63.54	60.76
		Panel C: ε_t is conditional t							
CM_n	100	4.54	5.16	4.98	4.94	8.88	7.38	10.08	9.26
	500	4.22	4.96	4.94	4.72	31.56	23.48	29.20	23.88
	1000	4.96	5.18	4.56	4.94	57.08	43.58	65.16	55.26
KS_n	100	4.66	4.66	5.14	5.24	8.50	7.28	11.58	10.52
	500	4.34	5.28	4.82	4.94	29.60	21.50	33.34	29.06
	1000	4.50	4.96	4.96	4.86	52.46	39.90	64.08	54.86
$\text{VR}_n(2)$	100	3.98	4.52	4.34	4.20	7.96	6.48	7.34	6.72
	500	4.42	5.12	5.16	4.78	32.24	24.02	8.32	7.38
	1000	5.44	4.98	4.90	4.96	56.98	44.00	8.76	8.60
$\text{VR}_n(5)$	100	4.68	4.36	4.58	4.92	7.62	6.80	11.20	10.02
	500	4.82	5.14	5.28	4.54	21.34	17.34	35.28	30.60
	1000	4.98	4.86	4.56	5.32	37.98	28.24	60.18	54.04
$\text{VR}_n(10)$	100	5.32	5.34	5.22	5.08	7.70	7.14	10.46	10.56
	500	4.54	4.72	4.80	4.88	13.34	10.98	30.08	25.82
	1000	5.50	4.44	4.50	5.28	22.30	18.06	52.24	46.80

power performance of $\mathrm{VR}_n(k)$ becomes worse as k increases, while when y_t exhibits the second order autocorrelation as in models 7-8, $\mathrm{VR}_n(2)$ has a very low power since it can only detect the first order autocorrelation, and the power of $\mathrm{VR}_n(5)$ is higher than that of $\mathrm{VR}_n(10)$. Fourth, CM_n and KS_n, taking into account of the autocorrelations at all lags, always have a competitive power performance with regard to the best performing $\mathrm{VR}_n(k)$. Overall, simulation studies indicate that both CM_n and KS_n have a good power performance with no risk of lag-selection as in $\mathrm{VR}_n(k)$ or size distortions.

4. Application

In this section, we apply the sign-based tests to several exchange rate series. The data sets we studied are the four daily currencies against the U.S. dollar, the Argentine Peso (USD/ARS), Chinese Yuan (USD/CNY), Colombian Peso (USD/COP), and Malaysian Ringgit (USD/MYR), over the period from November 14, 2009 to August 10, 2012. They are the currencies from the developing countries, two from Latin America and two from Asia. Each series has a total of 1001 observations. Denote the log-return ($\times 100$) of each series by $\{y_t\}_{t=1}^n$ with $n = 1000$. A simple visual inspection of the sample autocorrelation plots of $\{y_t^2\}_{t=1}^n$ in Figure 1 implies that all return series are highly correlated with possible ARCH effect. Moreover, Figure 2 plots the Hill's estimators $\{H_y(k)\}_{k=10}^{180}$ for each return series, where

$$H_y(k) = \left[\frac{1}{k}\sum_{i=1}^{k}\log\frac{y_{(n-i)}}{y_{(n-k)}}\right]^{-1}$$

with $\{y_{(t)}\}_{t=1}^n$ being the ascending order statistics of $\{y_t\}_{t=1}^n$. From Figure 2, it is reasonable to conclude that the tail indexes of USD/ARS, USD/COP, and USD/MYR return series are less than 4, and the tail index of USD/CNY return series is even less than 2. Thus, it is reasonable to use all sign-based tests to detect whether each return series is a MDS with conditional heteroscedasity.

Table 3 reports all of the results for each sign-based test. From this table, we find that (i) the tests CM_n, KS_n, and $\mathrm{VR}_n(5)$ imply that USD/ARS return series is not a MDS, while this can not be detected by others; (ii) all tests indicate that USD/CNY return series is a MDS; (iii) the tests CM_n, KS_n, and $\mathrm{VR}_n(2)$ have a very strong evidence to reject the MDS hypothesis for USD/COP return series, but this can not be detected by $\mathrm{VR}_n(5)$ or $\mathrm{VR}_n(10)$; and (iv) only KS_n can reject the MDS hypothesis for USD/MYR return series. Overall, our sign-based tests CM_n and KS_n give more consistent and much stronger rejections, while the results using the $\mathrm{VR}_n(k)$ tests are quite mixing. Interestingly, only the USD/CNY return series is a MDS from our test

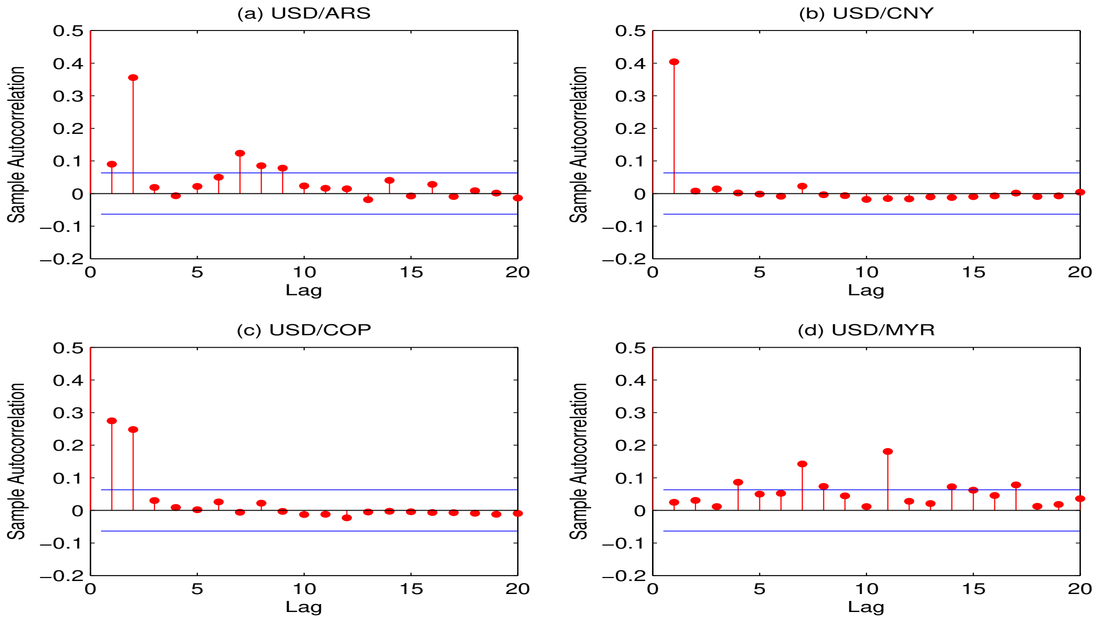

Fig. 1. Sample autocorrelation functions of $\{y_t^2\}$ for four different exchange rates.

results. It means that people in the exchange market can not get profit via predicting the value of USD/CNY. The reason is probably because CNY is not a freely convertible currency, and the USD/CNY is a managed floating exchange rate released by the People's Bank of China. For other aforementioned exchange rates, they do not have such a mechanism like USD/CNY, and people in these exchange markets can possibly conduct prediction by their own strategy.

Table 3. *The values of all sign-based tests.*

	Return series			
Tests	USD/ARS	USD/CNY	USD/COP	USD/MYR
CM_n	0.7617∗	0.3794	1.2382***	0.5413
KS_n	1.1132**	0.3974	1.8229***	0.9610∗
$\mathrm{VR}_n(2)$	-0.7589	1.8341	2.5931***	1.5179
$\mathrm{VR}_n(5)$	-2.0005∗	0.7939	0.5167	0.0318
$\mathrm{VR}_n(10)$	-0.9029	1.6521	1.2288	0.2323

[a] The test statistics have one, two, or three stars if significant at the level 5%, 2.5%, or 1%, respectively.

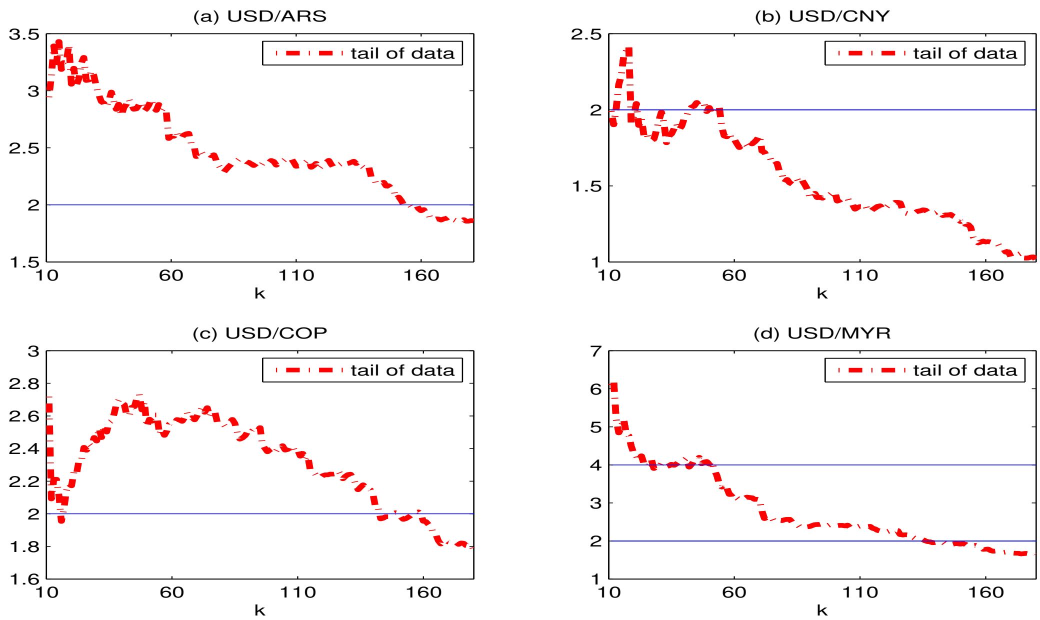

Fig. 2. Hill's estimators $\{H_y(k)\}_{k=10}^{180}$ for four different exchange rates.

5. CONCLUDING REMARKS

In this paper, we propose the sign-based CM and KS tests to detect the null hypothesis that the series is a MDS with conditional heteroscedasity. By only checking the autocorrelations of the signs, our new tests may not be consistent. However, as a compensation, our new tests allow for heavy-tailedness, non-stationarity, and nonlinear serial dependence of unknown forms. Particularly, they have the exact distributions, and hence no time-consuming bootstrap procedure is needed to obtain the critical values, and the size-distortion is not a problem any more. Generally, this is not the case for existing specification tests except the sign-based VR test in Wright (2000). Compared to Wright's (2000) test, our sign-based tests do not need to choose the lag parameter. This is indeed important, because simulation studies show that the power of sign-based VR test depends heavily on the choice of the lag, but our sign-based tests can always give a competitive power performance with regard to the best performing sign-based VR tests. Moreover, the empirical application shows that our sign-based CM and KS tests can give more consistent and much stronger rejections than the sign-based VR tests. Thus, in view of these, it is reasonable to recommend our sign-based CM and KS tests to practitioners.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (No.10990012, 11021161, 11371354, and 11201459).

REFERENCES

[1] BERKES, I. and HORVÁTH, L. (2004). The effciency of the estimators of the parameters in GARCH processes. *Annals of Statistics* **32**, 633–655.

[2] BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

[3] CHAN, K.-S., LI, D., LING, S. and TONG, H. (2014). On conditionally heteroscedastic AR models with thresholds. *Statistica Sinica* **24**, 625–652.

[4] CHEN, M. and ZHU, K. (2014). Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations. Forthcoming in *Journal of Econometrics*.

[5] CHEN, W. and DEO, R.S. (2006). The variance ratio statistic at large horizons. *Econometric Theory* **22**, 206–234.

[6] COUDIN, E. and DUFOUR, J.-M. (2009). Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *Econometrics Journal* **12**, S19–S49.

[7] DAVIS, R.A. and MIKOSCH, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* **26**, 2049–2080.

[8] DEO, R.S. (2000). Spectral tests of the martingale hypothesis under conditional heteroskedasticity. *Journal of Econometrics* **99**, 291–315.

[9] DURLAUF, S.N. (1991). Spectral-based testing of the martingale hypothesis. *Journal of Econometrics* **50**, 355–376.

[10] ENGLE, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica* **50**, 987–1008.

[11] ESCANCIANO, J.C. (2007). Model checks using residual marked emprirical processes. *Statistica Sinica* **17**, 115–138.

[12] FAN, J. and YAO, Q. (2003). Nonlinear time series: Nonparametric and parametric methods. Springer, New York.

[13] FRANCQ, C. and ZAKOÏAN, J.M. (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* **80**, 821–861.

[14] HALLIN, M., VERMANDELE, C., and WERKER, B. (2008). Semiparametrically efficient inference based on signs and ranks for median-restricted models. *Journal of the Royal Statistical Society, Series B* **70**, 389–412.

[15] HANSEN, B.E. (1994). Autoregressive conditional density estimation. *International Econoinic Review* **35**, 705–730.

[16] HONG, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.

[17] HONG, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association* **94**, 1201–1220.

[18] KOENKER, R. and BASSETT, G. (1982). Tests of linear hypotheses and L_1 estimation. *Econometrica* **50**, 1577–1584.

[19] LIM, K.-P. and BROOKS, R. (2011). The evolution of stock market efficiency over time: a survey of the empirical literature. *Journal of Economic Surveys* **25**, 69–108.

[20] LINTON, O., PAN, J. and WANG, H. (2010). Estimation for a non-stationary semi-strong GARCH$(1, 1)$ model with heavy-tailed errors. *Econometric Theory* **26**, 1–28.

[21] RACHEV, S.T. (2003). Handbook of Heavy Tailed Distributions in Finance. Elsevier/North-Holland.

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

[23] SHAO, X. (2011b). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27**, 312–343.

[24] SHEPHARD, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley, & O.E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and Other Fields*. Chapman and Hall.

[25] TIMMERMANN, A. and GRANGER, C.W.J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting* **20**, 15–27.

[26] TSAY, R.S. (2005). *Analysis of financial time series (2nd ed.).* New York: John Wiley & Sons, Incorporated.

[27] WRIGHT, J.H. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics* **18**, 1-9.

[28] ZHU, K. and LI, W.K. (2014). A bootstrapped spectral test for adequacy in weak ARMA models. Working paper. The university of Hong Kong.

[29] ZHU, K. and LING, S. (2014). Inference for ARMA models with unknown-form and heavy-tailed G/ARCH-type noises. Working paper. Hong Kong University of Science and Technology.